Mathematical Statistics I
Mar 2023 - Jun 2023
Tutoring

Tutor
Joonhyuk JUNG (정준혁)
Department of Statistics
25동 408호

1 Supplementary Material: Probability Space (확률 공간)

Definition 1. *Given a set S, the* ***power set (멱집합)*** *$\mathcal{P}(S)$ is the set of all subsets of S.*

For example, $\mathcal{P}(\emptyset) = \{\emptyset\}$.

Definition 2. *Given a nonempty set S, an* ***algebra (대수)*** *$\mathcal{F} \subset \mathcal{P}(S)$ is said to be a* ***σ-algebra*** *on S if it is closed under* ***countable union (가산 합집합)****, that is,*

- *$\emptyset \in \mathcal{F}$,*
- *$S \setminus A \in \mathcal{F}$ whenever $A \in \mathcal{F}$, and*
- *$\bigcup_{j=1}^{\infty} A_j \in \mathcal{F}$ whenever $A_j \in \mathcal{F}$ for $j \in \mathbb{N}$.*

Note that every σ-algebra is closed under **countable intersection (가산 교집합)**. The following are some examples.

- $\{\emptyset, S\}$ is the **trivial** σ-algebra on S if S is nonempty.
- $\mathcal{P}(S)$ is the **discrete (이산)** σ-algebra on S.
- $\{\emptyset, \{1,2\}, \{3\}, \{1,2,3\}\}$ is a σ-algebra on $\{1,2,3\}$.
- If S is uncountable, $\{A \in \mathcal{P}(S) : A \text{ or } S \setminus A \text{ is countable}\}$ is a σ-algebra on S.

Definition 3. *Given a* ***sample space (표본 공간)*** *S and a σ-algebra $\mathcal{F}$ on S, a member of $\mathcal{F}$ is called an* ***event (사건)****.*

Definition 4. *Given a σ-algebra $\mathcal{F}$, a nonnegative-real-valued function $\mu : \mathcal{F} \to [0, \infty)$ is said to be a* ***finite measure (유한 측도)*** *on $\mathcal{F}$ if*

- *$\mu(\emptyset) = 0$,*
- *$\mu\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mu(A_j)$ whenever $\{A_j\}_{j=1}^{\infty} \subset \mathcal{F}$ is a* ***disjoint (서로소)*** *sequence of members in $\mathcal{F}$.*

The second porperty is called the ***countable additivity of measure (가산가법성)****.*

Definition 5. *Given a sample space S and a σ-algebra $\mathcal{F}$ on S, a finite measure $\mathbb{P} : \mathcal{F} \to [0, \infty)$ on $\mathcal{F}$ is called a* ***probability measure (확률 측도)*** *on $\mathcal{F}$ if $\mathbb{P}(S) = 1$. The triple $(S, \mathcal{F}, \mathbb{P})$ is called a* ***probability space (확률 공간)****.*

Note that $0 \leq \mathbb{P}(A) \leq 1$ for all $A \in \mathcal{F}$.

Definition 6. *Given a probability space $(S, \mathcal{F}, \mathbb{P})$, a function $X : S \to \mathbb{R}$ is said to be a* ***random variable (확률 변수)*** *if it is $\mathcal{F}$-measurable, that is,*

- *$\{s \in S : X(s) \leq x\} \in \mathcal{F}$ for all $x \in \mathbb{R}$.*

Given X, one defines a function $F_X : \mathbb{R} \to [0,1]$ by $F_X(x) = \mathbb{P}(\{s \in S : X(s) \leq x\})$. It is called the ***cumulative distribution function (cdf; 누적 분포 함수)*** *of X. As a remark, $\mathbb{P}(X \leq x)$ is a shorthand form of the right hand side.*

- The cdf F_X of X is (1) non-decreasing, (2) right-continuous, and (3) satisfies $F_X(-\infty) = 0, F_X(\infty) = 1$.

2 Exercises a.k.a. 족보

- THERE IS NO ROYAL ROAD TO MATHEMATICAL STATISTICS.

2.1 BASIC QUESTION A

Let $\{A_j\}_{j=1}^{\infty}$ be a sequence of events. Prove that $\mathbb{P}\left(\liminf_{n\to\infty} A_n\right) \leq \liminf_{n\to\infty} \mathbb{P}(A_n)$ where

$$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{j=n}^{\infty} A_j$$

is an analogous definition of

$$\liminf_{n\to\infty} a_n = \lim_{n\to\infty} \inf \{a_j : j \geq n\}$$

in **해석개론 1**.

2.1.1 Answer

Define $B_n = \bigcap_{j=n}^{\infty} A_j$ for each $n \in \mathbb{N}$. One may show that $\{B_n\}_{n=1}^{\infty}$ is an increasing sequence of events. Now we are to show the inequality:

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} B_n\right) \leq \lim_{n\to\infty} \inf \{\mathbb{P}(A_j) : j \geq n\}.$$

By appealing to the continuity of probability measure in the textbook, $\mathbb{P}\left(\bigcup_{n=1}^{\infty} B_n\right) = \lim_{n\to\infty} \mathbb{P}(B_n)$ holds. Hence, it is enough to show that

$$\mathbb{P}(B_n) \leq \inf \{\mathbb{P}(A_j) : j \geq n\},$$

which is obvious from the monotonicity of probability measure in the textbook since $B_n \subset A_j$ for all $j \geq n$.

- ADDITIONAL NOTES

It is an easy exercise to show that

$$\mathbb{P}(\liminf A_n) \leq \liminf \mathbb{P}(A_n) \leq \limsup \mathbb{P}(A_n) \leq \mathbb{P}\left(\limsup A_n\right).$$

This inequalities are called the **continuity inequalities of measure** by some authors. In case $\liminf A_n = \limsup A_n$, one writes $\lim A_n = \liminf A_n = \limsup A_n$. (Otherwise, $\lim A_n$ is not defined.)
If $\lim A_n$ is well-defined, then

$$\begin{aligned} \mathbb{P}(\lim A_n) = \mathbb{P}(\liminf A_n) = \liminf \mathbb{P}(A_n) = \limsup \mathbb{P}(A_n) &= \mathbb{P}\left(\limsup A_n\right) \\ &= \lim \mathbb{P}(A_n). \end{aligned}$$

Note that $\lim A_n$ is well-defined when $\{A_n\}_{n=N}^{\infty}$ is either increasing or decreasing for some N. However, the converse is false. Consider for $S = \mathbb{N}$, $A_{2n} = \{1, \cdots, n\}$, $A_{2n-1} = \{1, \cdots, 2n-1\}$. Then $\lim A_n = \mathbb{N}$.

2.2 BASIC QUESTION B

Ler F be the cdf of a random variable X. Prove that $\mathbb{P}(X = x) = F(x) - F(x-)$ where

$$F(x-) = \lim_{y \uparrow x} F(y) = \lim_{h \downarrow 0} F(x-h)$$

2.2.1 Answer

Define $A_n = \{s \in S : X(s) \in (-\infty, x - \frac{1}{n}]\}$ for each $n \in \mathbb{N}$. Then $\{A_n\}_{n=1}^{\infty}$ is an increasing sequence of events. Hence by the continuity of probability measure,

$$\mathbb{P}(X < x) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n) = \lim_{n\to\infty} F\left(x - \frac{1}{n}\right) = F(x-).$$

By the additivity of probability measure, we have $\mathbb{P}(X = x) = \mathbb{P}(X \le x) - \mathbb{P}(X < x)$.

2.3 김우철 (2011)

Suppose $A_1, \cdots, A_n$ are events in the sample space S. Prove that

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \sum_{i<j<k} \mathbb{P}(A_i \cap A_j \cap A_k).$$

2.3.1 Answer

We further claim stronger proposition given by

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} \mathbb{P}(A_i) \tag{1}$$

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \ge \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{i<j\le n} \mathbb{P}(A_i \cap A_j) \tag{2}$$

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{i<j\le n} \mathbb{P}(A_i \cap A_j) + \sum_{i<j<k\le n} \mathbb{P}(A_i \cap A_j \cap A_k). \tag{3}$$

Proof by induction. It is easy to check the cases $n = 1, 2$. Firstly, we give a proof of the inequality (3) with respect to $A_1, \cdots, A_{n+1}$. Observe that

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{i=1}^{n+1} A_i\right) &= \mathbb{P}\left(\left(\bigcup_{i=1}^{n} A_i\right) \cup A_{n+1}\right) \\
&= \mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) + \mathbb{P}(A_{n+1}) - \mathbb{P}\left(\left(\bigcup_{i=1}^{n} A_i\right) \cap A_{n+1}\right) \\
&\le \sum_{i=1}^{n+1} \mathbb{P}(A_i) - \sum_{i<j\le n} \mathbb{P}(A_i \cap A_j) + \sum_{i<j<k\le n} \mathbb{P}(A_i \cap A_j \cap A_k) - \mathbb{P}\left(\bigcup_{i=1}^{n} (A_i \cap A_{n+1})\right).
\end{aligned}$$

This inequality holds by the induction hypothesis (3) applied to $A_1, \cdots, A_n$. Applying the induction hypothesis (2) to the last term regarding $A_1 \cap A_{n+1}, \cdots, A_n \cap A_{n+1}$ ends the proof. We omit proofs of (1), (2) for $n+1$ sets because they are much easier.

2.4 Unknown (2009) and 이재용 (2016)

Suppose $A_1, \cdots, A_n$ are events in the sample space S. Prove that

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \sum_{i<j<k} \mathbb{P}(A_i \cap A_j \cap A_k) - \cdots + (-1)^{n+1} \mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right)$$

2.4.1 Answer

Omitted.

2.5 Unknown (2009)

Events $A_1, \cdots, A_n$ in the sample space S is said to be "pairwise" independent, if

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j), i < j$$

Prove by a counter-example that the pairwise independence does not imply the independence of $A_1, \cdots, A_n$.

2.5.1 Answer

Given the sample space $S = \{0, 1, 2, 3\}$ and the collection of events $\mathcal{F} = \mathcal{P}(S)$, define the uniform probability as

$$\mathbb{P}(A) = \frac{|A|}{|S|} = \frac{|A|}{4}.$$

To put it more intuitive and simple, just roll a fair regular-tetrahedral die. **정사면체 주사위를 하나 굴린다.** Fix $n = 3$ and define $A_i = \{0, i\}$ for $i = 1, 2, 3$. Confirm that the events A_1, A_2, A_3 are pairwise independet but not "mutually" independent.

2.6 이재용 (2020)

Let (X, Y, Z) be jointly distributed with the pdf

$$f(x, y, z) = \frac{1 - \sin x \sin y \sin z}{8\pi^3} \mathrm{I}(0 \le x, y, z \le 2\pi).$$

Prove that X, Y, Z are pairwise independent, but not independent as a 3-dimensional random vector.

2.6.1 Answer

Integrating out z gives the joint pdf of (X, Y):

$$\begin{aligned} f_{1,2}(x,y) &= \int_0^{2\pi} \frac{1-\sin x \sin y \sin z}{8\pi^3} dz \mathrm{I}(0 \le x, y \le 2\pi) \\ &= \frac{1}{4\pi^2}\mathrm{I}(0 \le x, y \le 2\pi) \\ &= \frac{1}{2\pi}\mathrm{I}(0 \le x \le 2\pi)\frac{1}{2\pi}\mathrm{I}(0 \le y \le 2\pi) \\ &= f_1(x)f_2(y). \end{aligned}$$

Hence X, Y are independent. However, it is obvious that $f_{1,2,3}(x,y,z) \neq f_1(x)f_2(y)f_3(z)$.

2.7 Unknown (2007*, 2009*) and 김우철 (2015*, 2017)

Suppose the cdf F of X is given by

$$F(x) = \begin{cases} 0, & x < 0 \\ (x^2+1)/9, & 0 \le x < 1 \\ (x^2+4)/9, & 1 \le x < 2 \\ 1, & x \ge 2 \end{cases}$$

For $k = 1, 2, \cdots$, define $A_k = [1/k, 2 - 1/k]$ and $B_k = (1 - 1/k, 2 + 1/k)$. Find the following: $\lim_{k\to\infty} A_k, \lim_{k\to\infty} \mathbb{P}(X \in A_k), \lim_{k\to\infty} B_k, \lim_{k\to\infty} \mathbb{P}(X \in B_k)$.

2.7.1 Answer

Note that A_k is increasing and B_k is decreasing so that $\lim_{k\to\infty} A_k$ and $\lim_{k\to\infty} B_k$ are well-defined. Verify that

$$\begin{aligned} &\lim_{k\to\infty} A_k = (0,2) && \lim_{k\to\infty} \mathbb{P}(X \in A_k) = F(2-) - F(0) = 8/9 - 1/9 = 7/9 \\ &\lim_{k\to\infty} B_k = [1,2] && \lim_{k\to\infty} \mathbb{P}(X \in B_k) = F(2) - F(1-) = 1 - 2/9 = 7/9 \end{aligned}$$

2.8 김우철 (2018)

Suppose the cdf F of X is given by

$$F(x) = \begin{cases} 0, & x < 0 \\ x/10, & 0 \le x < 2 \\ x^2/10, & 2 \le x < 3 \\ 1, & x \ge 4 \end{cases}$$

For $n = 1, 2, \cdots$, define $B_n = (2 - 1/n, 3 - 1/n)$. Prove that $\liminf_{n\to\infty} B_n = \limsup_{n\to\infty} B_n$ and find $\mathbb{P}(X \in \lim_{n\to\infty} B_n)$.

2.8.1 Answer

Note that B_n is neither increasing nor decreasing. However,

$$\liminf B_n = \bigcup_{n=1}^{\infty} \bigcap_{j=n}^{\infty} B_j = \bigcup_{n=1}^{\infty} \left[2, 3 - \frac{1}{n}\right) = [2, 3)$$

and

$$\limsup B_n = \bigcap_{n=1}^{\infty} \bigcup_{j=n}^{\infty} B_j = \bigcap_{n=1}^{\infty} \left(2 - \frac{1}{n}, 3\right) = [2, 3)$$

coincide. Therefore, $\lim B_n = [2, 3)$ and hence

$$\mathbb{P}(X \in \lim B_n) = F(3-) - F(2-) = 9/10 - 2/10 = 7/10.$$

2.9 이재용 (2016)

Let F be the cdf of X. Define $F^{-1}(u) = \inf\{x \in \mathbb{R} : F(x) \geq u\}$ for $0 < u < 1$. Prove the following:

- F^{-1} is well-defined. (Why does the infimum exist?)
- $F(F^{-1}(u)) \geq u$ for $u \in (0, 1)$.
- $F^{-1}(F(x)) \leq x$ for $x \in \mathbb{R}$.
- $u \leq F(x) \iff F^{-1}(u) \leq x$ for $u \in (0, 1)$ and $x \in \mathbb{R}$.
- Suppose F is continuous **and strictly increasing**. Then $F(F^{-1}(u)) = u$ and $F^{-1}(F(x)) = x$.

2.9.1 Answer

Given $u \in (0, 1)$, write $A_u = \{x \in \mathbb{R} : F(x) \geq u\}$. Verify that A_u is a nonempty subset of $\mathbb{R}$ bounded from below. Hence $F^{-1}(u) = \inf A_u$ is well-defined. For each $n \in \mathbb{N}$, there exists an element $x_n \in A_u$ such that

$$x_n < \inf A_u + \frac{1}{n} = F^{-1}(u) + \frac{1}{n}.$$

Since F is non-decreasing, one has

$$u \leq F(x_n) \leq F\left(F^{-1}(u) + \frac{1}{n}\right).$$

Taking $\lim_{n\to\infty}$ concludes that $u \leq F(F^{-1}(u))$ since F is right-continuous. Now for each $x \in \mathbb{R}$, it is obvious that

$$x \in A_{F(x)},$$

implying that $x \geq \inf A_{F(x)} = F^{-1}(F(x))$. Now we are to show $u \leq F(x) \iff F^{-1}(u) \leq x$. Since F is non-decreasing, it only remains to elaborate that F^{-1} is non-decreasing (left to the tutees).

Suppose now F is continuous and strictly increasing. It is easy to check that $F(\mathbb{R}) = (0,1)$. That is, for each $u \in (0,1)$, there exists $x \in \mathbb{R}$ such that $F(x) = u$. Conversely, for each $x \in \mathbb{R}$, set $u = F(x) \in (0,1)$. Observe that $A_u = [x, \infty)$. As a result,

$$F^{-1}(u) = \inf A_u = \inf[x, \infty) = x.$$

Explain why this ends the proof.

2.10 Unknown (2007, 2009, 2011)

Let X be a non-negative random variable of continuous type with pdf f and cdf F satisfying $F'(x) = f(x)$ for all $x > 0$. Suppose $\mathbb{E}(X) < \infty$. Prove that $\lim_{x\to\infty} x(1 - F(x)) = 0$ and $\mathbb{E}(X) = \int_0^\infty (1 - F(x))dx$.

2.10.1 Answer

Given a constant $x > 0$,

$$0 \le x(1 - F(x)) = x \int_x^\infty f(z)dz \le \int_x^\infty zf(z)dz = \int_0^\infty zf(z)dz - \int_0^x zf(z)dz.$$

This argument is valid since $zf(z)$ is nonnegative for $z > 0$ and $\int_0^\infty zf(z)dz = \mathbb{E}(X) < \infty$ by assumption. The right hand side converges to zero as $x \to \infty$. As a result, $\lim_{x\to\infty} x(1 - F(x)) = 0$ by appealing to the Sandwich Theorem. Indeed, Fubini's Theorem applied to a nonnegative function ensures us that

$$\mathbb{E}(X) = \int_0^\infty zf(z)dz = \int_0^\infty \int_0^z f(z)dxdz = \int_0^\infty \int_x^\infty f(z)dzdx = \int_0^\infty (1 - F(x))dx.$$

2.11 Unknown (2007)

Suppose that X and Y have the joint pdf

$$f_{1,2}(x, y) = 15x^2 y \mathrm{I}(0 < x < y < 1).$$

Compute $\mathbb{P}(Y \le 1/2)$ and $\mathbb{P}(X + Y \le 1)$.

2.11.1 Answer

$$
\begin{aligned}
\mathbb{P}(Y \le 1/2) &= \int_0^{1/2} \int_0^{y} 15x^2 y \, dxdy \\
&= \int_0^{1/2} 5y^4 \, dy = (1/2)^5 = 1/32 \\
\mathbb{P}(X + Y \le 1) &= \int_0^{1/2} \int_x^{1-x} 15x^2 y \, dydx \\
&= \int_0^{1/2} \frac{15}{2} x^2 (1 - 2x) \, dx \\
&= \int_0^{1} \frac{15}{16} z^2 (1 - z) \, dz && \text{(substitute } z = 2x\text{)} \\
&= \frac{15}{16} \cdot \frac{1}{12} = \frac{5}{64}.
\end{aligned}
$$

2.12 Unknown (2007)

The pdf of standard logistic distribution $L(0, 1)$ is given by

$$f(x) = \frac{e^{-x}}{(1 + e^{-x})^2} \mathrm{I}(-\infty < x < \infty).$$

Find $F^{-1}(u)$ for the cdf F of $L(0, 1)$.

2.12.1 Answer

Define $\sigma(x) = 1/(1 + e^{-x})$ for $x \in \mathbb{R}$. Verify that $\sigma \in C^\infty(\mathbb{R})$, that is, σ is k times continuously differentiable for each $k \in \mathbb{N}$.

$$
\begin{aligned}
\sigma(-x) &= 1 - \sigma(x) \\
\sigma'(x) &= \sigma(x)\sigma(-x) = f(x) \\
\sigma^{-1}(u) &= \log \frac{u}{1 - u}
\end{aligned}
$$

Then $F(x) = \int_{-\infty}^{x} \sigma'(z) dz = \sigma(x)$ and hence $F^{-1}(u) = \log \frac{u}{1-u}$. How can you get a random sample from the standard logistic distribution? Consider $X = \log \frac{U}{1-U}$ where $U \sim$ the standard uniform distribution.

2.13 Unknown (2011)

Let (X, Y) be jointly distributed with the pdf

$$f(x, y) = y^{-1} e^{-y} \mathrm{I}(0 < x < y < \infty).$$

Find the marginal pdf $f_1(x)$, the conditional pdf $f_{2|1}(y|x)$, and $\mathrm{Var}[Y|X]$.

2.13.1 Answer

$$
\begin{aligned}
f_1(x) &= \int_x^\infty y^{-1}e^{-y}\,dy && \text{(non-elementary; no closed form)}\\
f_{2|1}(y|x) &= \frac{y^{-1}e^{-y}\mathrm{I}(x<y<\infty)}{f_1(x)}\\
\mathbb{E}[Y|X=x] &= \frac{\int_x^\infty e^{-y}\,dy}{f_1(x)}\\
&= \frac{e^{-x}}{f_1(x)}\\
\mathbb{E}[Y^2|X=x] &= \frac{\int_x^\infty ye^{-y}\,dy}{f_1(x)}\\
&= \frac{(1+x)e^{-x}}{f_1(x)}
\end{aligned}
$$

Recall that

$$\mathrm{Var}[Y|X=x] = \mathbb{E}[Y^2|X=x] - (\mathbb{E}[Y|X=x])^2$$

and hence

$$\mathrm{Var}(Y|X) = \frac{(1+X)e^{-X}}{f_1(X)} - \left(\frac{e^{-X}}{f_1(X)}\right)^2$$

where

$$f_1(X) = \int_X^\infty y^{-1}e^{-y}dy.$$

2.14 Unknown (2009)

Let (X,Y) be jointly distributed with the pdf

$$f(x,y) = \frac{1+xy}{4}\mathrm{I}(|x|<1)\mathrm{I}(|y|<1).$$

Prove that X and Y are NOT independent. Prove that X^2 and Y^2 are independent.

2.14.1 Answer

Hint: Compute $\mathbb{P}(X^2<t)$ given $0<t<1$. Details are left to the tutees.

2.15 Unknown (2009)

Let F be the cdf of a random variable. Prove that the set of discontinuity points of F,

$$\mathcal{D} = \{x\in\mathbb{R} : F \text{ is discontinuous at } x\}$$

is countable.

2.15.1 Answer

There exists a function $h : \mathcal{D} \to \mathbb{Q}$ such that

$$h(x) \text{ is an arbitary element of } (F(x-), F(x)) \cap \mathbb{Q}.$$

h is well-defined since $F(x-) < F(x)$ for every $x \in \mathcal{D}$ and $\mathbb{Q}$ is dense in $\mathbb{R}$.
(Remark: This argument depends on the **Axiom of Choice**. Search it if you are interested.)
For $x < y$ in $\mathcal{D}$, we have

$$F(x-) < F(x) \leq F(y-) < F(y)$$

because $y_n := y - \frac{1}{n}$ converges to y from below and there exists $N \in \mathbb{N}$ such that

$$n > N \Longrightarrow x < y_n \Longrightarrow F(x) \leq F(y_n).$$

As a result, h is injective, implying that $\mathcal{D}$ is countable.

3 Remark

Definition 7. *A random variable $X : S \to \mathbb{R}$ is said to be **of discrete type (이산형 확률 변수)** if the image*

$$X(S) = \{X(s) \in \mathbb{R} : s \in S\}$$

*is **discrete** in the sense that every point in $X(S)$ is **isolated (고립점)**.*

Recall that a point p is said to be an **isolated** point of a subset A in the metric space $\mathbb{R}$ if there exists an open neighborhood of p that does not contain any other points of A.

Regardless of the geometry of the image $X(S)$, the cdf F_X of X is defined on the entire real line $\mathbb{R}$.

Definition 8. *A random variable $X : S \to \mathbb{R}$ is said to be **of continuous type (연속형 확률 변수)** if the cdf F_X of X is continuous.*

Definition 9. *A random variable $X : S \to \mathbb{R}$ is said to be **of mixed type (혼합형 확률 변수)** if it is neither discrete nor continuous, but is a mixture of both.*

4 Advanced Exercise

Suppose the cdf F of a random variable X is given by

$$F(x) = \begin{cases} a\arctan(x) + \pi a + b, & x \geq 0 \\ \\ a\arctan(x) + b, & x < 0 \end{cases}$$

for some constants $a, b \in \mathbb{R}$.

- Find a and b.
- **(optional)** Articulate that $(X \in \mathbb{Q})$ and $(X < 1)$ are events.
- Compute $\mathbb{P}(X \in \mathbb{Q})$ and $\mathbb{P}(X < 1)$.
- Verify that the random variable X is neither discrete nor continuous.
- Does a pdf f of X exist?
- Prove that there exist random variables X_d and X_c such that X_d is discrete, X_c is continuous, and

  $$F = \lambda F_{X_d} + (1 - \lambda) F_{X_c}$$

  holds for some $\lambda \in [0, 1]$.
- **(optional)** How would you generate a random sample $X \sim F$ in practice?

1 Supplementary Material

1.1 Infinite Sum

Definition 1. *Given an infinite set A and function $f : A \to [0, \infty)$, the* ***infinite sum*** $\sum_{a \in A} f(a)$ *is defined by*

$$\sum_{a \in A} f(a) = \sup \left\{ \sum_{b \in B} f(b) : B \subseteq A, B \text{ is finite} \right\}.$$

Proposition 1. *Suppose A is a* ***countably*** *infinite set. Then there exists a* ***bijection (전단사 함수)***

$$r : \mathbb{N} \to A.$$

Proposition 2. *Given a* ***countably*** *infinite set A and function $f : A \to [0, \infty)$,*

$$\sum_{a \in A} f(a) = \sum_{n=1}^{\infty} f(r(n))$$

holds for every bijection $r : \mathbb{N} \to A$.

2.4 급수의 수렴판정

이 장에서는 급수의 수렴판정법을 몇 가지 공부하는데, 그 주요 방법은 코시 판정법이다. 급수의 수렴판정을 논하기 앞서서, 급수의 항들을 새로이 결합하거나 교환하는 경우 어떤 일이 일어나는지 살펴보자. 예를 들어서 급수 $\sum a_n$ 을

$$(a_1 + a_2) + a_3 + (a_4 + a_5 + a_6) + (a_7 + a_8) + \dots$$

로 다시 결합하자. 급수 $\sum a_n$ 의 부분합의 수열을 $\langle s_n \rangle$ 이라 두면, 위 새 급수의 합은 수열 $\langle s_2, s_3, s_6, s_8, \dots \rangle$ 의 극한이 어떻게 되는가 하는 문제로 귀결된다. 따라서, 만일 급수 $\sum a_n$ 이 수렴한다면 이를 다른 방법으로 결합하여도 수렴 여부나 급수의 합은 변하지 않는다. 물론, 수렴하지 않는 급수의 경우는 함부로 괄호를 칠 수 없다는 것을 잘 알고 있을 것이다.

이제, 급수의 항을 교환하는 문제를 살펴보자. 급수 $\sum_n a_n$ 과 전단사함수 $r : \mathbb{N} \to \mathbb{N}$ 이 주어져 있을 때, 급수 $\sum_n a_{r(n)}$ 을 $\sum_n a_n$ 의 재배열급수라 한다.

명제 2.4.1. 각 $n = 1, 2, \dots$ 에 대하여 $a_n \geq 0$ 이고, 함수 $r : \mathbb{N} \to \mathbb{N}$ 이 전단사 함수라 하자. 만일 급수 $\sum_n a_n$ 이 s 로 수렴하면, $\sum_n a_{r(n)} = s$ 이다. 또한, $\sum_n a_n$ 이 발산하면 $\sum_n a_{r(n)}$ 역시 발산한다.

1.2 Joint Cumulant Generating Function

Recall that the **joint moment generating function (결합적률생성함수)** of X, Y is defined as

$$M_{1,2}(t_1, t_2) = \mathbb{E}\left(e^{t_1 X + t_2 Y}\right)$$

if the expectation is finite in some open neighborhood of $(t_1, t_2) = (0, 0)$.

Proposition 3. *If mgf $M_{1,2}(t_1, t_2)$ of (X, Y) exists, (i.e., $\mathbb{E}(e^{t_1 X + t_2 Y}) < \infty$ for (t_1, t_2) contained in an open neighborhood of the origin) then the joint moments $\mathbb{E}(X^i Y^j)$ of all orders are well-defined. In addition,*

$$M_{1,2}(t_1, t_2) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{\mathbb{E}(X^i Y^j)}{i! j!} t_1^i t_2^j$$

holds in some open neighborhood of the origin. On the right hand side, we assume $0^0 = 1$ by convention.

The first several terms are very useful.

$$M_{1,2}(t_1, t_2) = 1 + (\mathbb{E}(X)t_1 + \mathbb{E}(Y)t_2) + \left(\mathbb{E}(X^2)\frac{t_1^2}{2} + \mathbb{E}(XY)t_1 t_2 + \mathbb{E}(Y^2)\frac{t_2^2}{2}\right) + O(\|t\|^3)$$

The natural logarithm of joint mgf is called **joint cumulant generating function (결합누율생성함수)**.

$$C_{1,2}(t_1, t_2) = \log M_{1,2}(t_1, t_2) = \log \mathbb{E}\left(e^{t_1 X + t_2 Y}\right)$$

The first several terms are attained from the series expansion: $\log(1 + x) = x - \frac{x^2}{2} + O(x^3)$.

$$C_{1,2}(t_1, t_2) = (\mathbb{E}(X)t_1 + \mathbb{E}(Y)t_2) + \left(\mathrm{Var}(X^2)\frac{t_1^2}{2} + \mathrm{Cov}(X, Y)t_1 t_2 + \mathrm{Var}(Y^2)\frac{t_2^2}{2}\right) + O(\|t\|^3)$$

1.3 Gamma Integral

Definition 2. *Gamma function $\Gamma : (0, \infty) \to \mathbb{R}$ is defined by*

$$\Gamma(t) = \int_0^{\infty} x^{t-1} e^{-x} \, dx$$

It is easily derived that

- $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$
- $\Gamma(1) = 1$
- $\Gamma(t + 1) = t\Gamma(t)$ for all $t > 0$. Henceforth, $\Gamma(n) = (n - 1)!$ for $n \in \mathbb{N}$. Note that $0! = 1$.

It is helpful to memorize that

$$\int_0^{\infty} x^{n-1} e^{-ax} \, dx = \frac{1}{a^n} \int_0^{\infty} z^{n-1} e^{-z} \, dz = \frac{\Gamma(n)}{a^n} = \frac{(n-1)!}{a^n}$$

holds for all $n \in \mathbb{N}$ and $a > 0$.

2 Exercises a.k.a. 족보

- THERE IS NO ROYAL ROAD TO MATHEMATICAL STATISTICS.

2.1 Advanced Exercise

Suppose the cdf F of a random variable X is given by

$$F(x) = \begin{cases} a\arctan(x) + \pi a + b, & x \geq 0 \\ \\ a\arctan(x) + b, & x < 0 \end{cases}$$

for some constants $a, b \in \mathbb{R}$.

- Find a and b.
- **(optional)** Articulate that $(X \in \mathbb{Q})$ and $(X < 1)$ are events.
- Compute $\mathbb{P}(X \in \mathbb{Q})$ and $\mathbb{P}(X < 1)$.
- Verify that the random variable X is neither discrete nor continuous.
- Does a pdf f of X exist?
- Prove that there exist random variables X_d and X_c such that X_d is discrete, X_c is continuous, and

  $$F = \lambda F_{X_d} + (1-\lambda)F_{X_c}$$

  holds for some $\lambda \in [0, 1]$.
- **(optional)** How would you generate a random sample $X \sim F$ in practice?

2.1.1 Answer

Solving $\lim_{x\to\infty} F(x) = 1$ and $\lim_{x\to-\infty} F(x) = 0$ gives

$$a = \frac{1}{2\pi} \qquad\qquad b = \frac{1}{4}$$

Recall that $(X \leq x)$ is an event for every $x \in \mathbb{R}$. Indeedly,

$$(X \in \mathbb{Q}) = \bigcup_{q\in\mathbb{Q}} (X = q) = \bigcup_{q\in\mathbb{Q}} \bigcap_{n=1}^{\infty} \Big((X \leq q) \setminus (X \leq q - \frac{1}{n}) \Big)$$
$$(X < 1) = \bigcup_{n=1}^{\infty} (X \leq 1 - \frac{1}{n})$$

are events since $\mathbb{Q}$ is countable. By appealing to the countable additivity and continuity of $\mathbb{P}$, we have

$$\begin{aligned}
\mathbb{P}(X \in \mathbb{Q}) &= \sum_{q \in \mathbb{Q}} \mathbb{P}\left(\bigcap_{n=1}^{\infty}\left((X \le q) \setminus (X \le q - \frac{1}{n})\right)\right) \\
&= \sum_{q \in \mathbb{Q}} \lim_{n\to\infty} \mathbb{P}\left((X \le q) \setminus (X \le q - \frac{1}{n})\right) \\
&= \sum_{q \in \mathbb{Q}} \lim_{n\to\infty} \left(F(q) - F(q - \frac{1}{n})\right) \\
&= \sum_{q \in \mathbb{Q}} (F(q) - F(q-)) \\
&= F(0) - F(0-) = \frac{1}{2}
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{P}(X < 1) &= \lim_{n\to\infty} \mathbb{P}(X \le 1 - \frac{1}{n}) \\
&= \lim_{n\to\infty} F(1 - \frac{1}{n}) \\
&= F(1-) = \frac{7}{8}
\end{aligned}$$

F is strictly increasing for an open interval (e.g., for $1 < x < 2$) so F is not discrete. However, F is discontinuous at $x = 0$, implying that F is not continuous as well. In addition, pdf f of X cannot be defined since F is discontinuous at $x = 0$. Now consider the following two cdfs.

$$F_{X_d}(x) = \begin{cases} 1, & x \ge 0 \\ 0, & x < 0 \end{cases} \qquad\qquad F_{X_c}(x) = \frac{1}{\pi}\arctan(x) + \frac{1}{2}$$

Verify that these are well-defined cdfs of discrete and continuous random variables, respectively. Furthermore, one has $F = \lambda F_{X_d} + (1-\lambda)F_{X_c}$ with $\lambda = 1/2$. Consider the following **hierarchical** random variable.

$$\begin{aligned}
Y &\sim \text{Ber}(\lambda) \\
X|Y = 1 &\sim F_{X_d} \\
X|Y = 0 &\sim F_{X_c}
\end{aligned}$$

Then it's an easy exercise to show that X exactly has F as its cdf. As described in the class, one can also prove that

$$F_{X_c}^{-1}(U) = \tan\left(\pi(U - \frac{1}{2})\right) \sim F_{X_c}$$

where $U \sim \text{Unif}(0,1)$ (i.e, the standard uniform distribution). Now we are able to generate n iid samples from the distribution F following the notion of mixed-type distribution. To put it more precise, given n iid standard uniform samples $U_1, \cdots, U_n$, compute $X_i = \tan\left(\pi(U_i - \frac{1}{2})\right)$ and coerce it into zero with probability 1/2 for each i.

2.2 이재용 (2016)

Let X be a continuous random variable endowed with the pdf f_X given by

$$f_X(x) = \begin{cases} 2x, & 0 \le x \le 1 \\ 0, & otherwise \end{cases}$$

Define Y by

$$Y = g(X) = \begin{cases} X, & 0 \le X \le \frac{1}{2} \\ \frac{1}{2}, & X > \frac{1}{2} \end{cases}$$

(a) Compute the cdf of Y.
(b) Compute the conditional pdf of Y given $Y < \frac{1}{2}$.

2.2.1 Answer

(a) Fix $0 \le y < \frac{1}{2}$. Then

$$\begin{aligned} \mathbb{P}(Y \le y) &= \mathbb{P}(0 \le X \le \frac{1}{2})\mathbb{P}(Y \le y | 0 \le X \le \frac{1}{2}) + \mathbb{P}(X > \frac{1}{2})\underbrace{\mathbb{P}(Y \le y | X > \frac{1}{2})}_{=0} \\ &= \mathbb{P}(0 \le X \le \frac{1}{2}, Y \le y) \\ &= \mathbb{P}(0 \le X \le y) = y^2 \end{aligned}$$

and hence

$$F_Y(y) = \mathbb{P}(Y \le y) = \begin{cases} 0, & y < 0 \\ y^2, & 0 \le y < \frac{1}{2} \\ 1, & y \ge \frac{1}{2} \end{cases}$$

(b) Note that $Y < \frac{1}{2}$ if and only if $0 \le X < \frac{1}{2}$. In particular, $Y = X$ holds provided that $Y < \frac{1}{2}$. Hence it only remains to compute the conditional pdf of X given $X < \frac{1}{2}$.

$$f_{Y|Y<\frac{1}{2}}(y) = 8y\mathrm{I}(0 \le y < \frac{1}{2})$$

2.3 이재용 (2016)

Consider a bivariate random variable (X, Y) with

$$\mathbb{E}(X) = \mu_1, \quad \mathrm{Var}(X) = \sigma_1^2, \quad \mathbb{E}(Y) = \mu_2, \quad \mathrm{Var}(Y) = \sigma_2^2, \quad \mathrm{Corr}(X, Y) = \rho.$$

Suppose all the qunatities are finite. Suppose $\mathbb{E}(Y|X) = a + bX$ for some reals $a, b \in \mathbb{R}$.
(a) Prove that $\mathbb{E}(Y|X) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(X - \mu_1)$.
(b) Prove that $\mathbb{E}(\mathrm{Var}(Y|X)) = \sigma_2^2(1 - \rho^2)$.

2.3.1 Answer

By appealing to the law of iterated expectations, we have

$$\mu_2 = \mathbb{E}(Y) = \mathbb{E}\left(\mathbb{E}(Y|X)\right) = \mathbb{E}(a+bX) = a + b\mu_1$$
$$\mu_1\mu_2 + \rho\sigma_1\sigma_2 = \mathbb{E}(XY) = \mathbb{E}\left(\mathbb{E}(XY|X)\right) = \mathbb{E}\left(X\mathbb{E}(Y|X)\right) = \mathbb{E}(aX + bX^2) = a\mu_1 + b(\mu_1^2 + \sigma_1^2)$$

Combining these two equations concludes that

$$a = \mu_2 - b\mu_1 \qquad\qquad b = \rho\frac{\sigma_2}{\sigma_1}$$

By appealing to the law of total variance,

$$\mathbb{E}\left(\mathrm{Var}(Y|X)\right) = \mathrm{Var}(Y) - \mathrm{Var}\left(\mathbb{E}(Y|X)\right) = \sigma_2^2 - b^2\sigma_1^2 = \sigma_2^2(1-\rho^2).$$

2.4 김우철 (2017)

Suppose a random variable X has its cgf (cumulant generating function). The r-th cumulant is given by

$$c_r = \begin{cases} (2k-1)!2^{-2k+1}, & r = 2k \\ 0, & r = 2k-1 \end{cases}$$

for $k = 1, 2, \cdots$.
(a) Find the r-th moment of X.
(b) Find the pdf of X.
(c) Find the kurtosis (첨예도) of X.

2.4.1 Answer

Consider its cgf $C_X(t)$.

$$\begin{aligned} C_X(t) &= \sum_{k=1}^{\infty} c_{2k}\frac{t^{2k}}{(2k)!} \\ &= \sum_{k=1}^{\infty}(2k-1)!2^{-2k+1}\frac{t^{2k}}{(2k)!} \\ &= 2\sum_{k=1}^{\infty}\frac{(t/2)^{2k}}{2k} \\ &= \sum_{n=1}^{\infty}\frac{(t/2)^n}{n} + \sum_{n=1}^{\infty}\frac{(-t/2)^n}{n} && \text{(All terms of odd indices cancel out.)} \\ &= -\log(1-\frac{t}{2}) - \log(1+\frac{t}{2}) && (-\log(1-x) = x + x^2/2 + x^3/3 + \cdots) \end{aligned}$$

whose radius of convergence is 2. Hence the mgf M_X of X is attained as follows.

$$M_X(t) = \exp C_X(t) = \left(1 - \frac{t^2}{4}\right)^{-1} = 1 + \sum_{k=1}^{\infty} \left(\frac{t^2}{4}\right)^k \qquad ((1-x)^{-1} = 1 + x + x^2 + x^3 + \cdots)$$
$$= 1 + \sum_{k=1}^{\infty} (2k)! 2^{-2k} \frac{t^{2k}}{(2k)!}$$

That is, the r-th moment is given by

$$m_r = \begin{cases} (2k)!2^{-2k}, & r = 2k \\ 0, & r = 2k-1 \end{cases}$$

for $k = 1, 2, \cdots$. Now we are to find the pdf of X. Recall that $-\log(1-t/2)$ is exactly the cgf of $Y \sim \mathrm{Exp}(1/2)$ endowed with the pdf

$$f_Y(y) = 2e^{-2y} \mathrm{I}(y > 0).$$

In fact, $-\log(1 + t/2)$ is nothing but the cgf of $-Z$ where $Z \sim \mathrm{Exp}(1/2)$. Hence, the Theorem 2.5.11(b) in the textbook says that $C_X(t)$ is explicitly the cgf of $Y - Z$ where Y, Z are iid(i.e, identical and independent). By the uniqueness of cgf illustated in the Theorem 2.2.4(b), we are ensured to write that

$$X \stackrel{d}{=} Y - Z$$

It is intriguing to show the following property of Exponential distributions, namely, the **memoryless property**.

$$(Y - Z | Y > Z) \stackrel{d}{=} Y$$
$$(Z - Y | Y \le Z) \stackrel{d}{=} Z$$

Let F denote the cdf of X. Given $x \ge 0$, we have

$$1 - F(x) = \mathbb{P}(X > x) = \underbrace{\mathbb{P}(Y > Z)}_{=1/2} \underbrace{\mathbb{P}(Y - Z > x | Y > Z)}_{=\mathbb{P}(Y > x) = e^{-2x}} = \frac{1}{2} e^{-2x}$$

Analogously, given $x \ge 0$,

$$F(-x) = \mathbb{P}(X \le -x) = \mathbb{P}(Y \le Z)\mathbb{P}(Z - Y \ge x | Y \le Z) = \frac{1}{2} e^{-2x}$$

Combining these two equations uniquely and entirely determines the values of F for all $x \in \mathbb{R}$. Now F is continuous and one has

$$f(x) = \frac{d}{dx} F(x) = e^{-2|x|} \mathrm{I}(-\infty < x < \infty).$$

The distribution F is called the **Double Exponential (이중지수)** distribution, a.k.a. **Laplace** distribution (with **location parameter (위치모수)** 0 and **scale parameter (척도모수)** $1/2$.) Please refer to https://en.wikipedia.org/wiki/Laplace_distribution. Then see Exercises 1.15 and 1.16 in the textbook. Finally, from the formula of the r-th cumumant, we have

$$\begin{aligned} c_2 &= 1!2^{-1} = 1/2, && (= \mathrm{Var}(X)) \\ c_4 &= 3!2^{-3} = 3/4. \end{aligned}$$

Then Exercises 1.19 and 1.20 say that the (excess) kurtosis of X equals to c_4/c_2^2.

$$\mathrm{kurt}(X) = \frac{c_4}{c_2^2} = 3.$$

That is, the Laplace distribution is much **sharper** than the normal distribution. Now see Exercise 1.22. As a final remark, kurtosis is **translation/scaling invariant** by definition. That is, for example,

- Every Normal distribution has 0 as its kurtosis.
- Every Laplace distribution has 3 as its kurtosis.

2.5 김우철 (2017)

Suppose X, Y are jointly distributed with the following pdf.

$$f_{1,2}(x, y) = 3e^{-2x-y}\mathrm{I}(0 < x < y < \infty)$$

(a) Find the conditional pdf $f_{2|1}(y|x)$ given $X = x$ for some $x > 0$.
(b) Find $\mathrm{Var}[\mathbb{E}(Y|X)]$ and $\mathbb{E}[\mathrm{Var}(Y|X)]$.
(c) Find $\mathrm{Var}[X + Y - \mathbb{E}(Y|X)]$.

2.5.1 Answer

A little calculus ensures us that

$$\begin{aligned} f_1(x) &= 3e^{-3x}\mathrm{I}(0 < x < \infty) \\ f_{2|1}(y|x) &= e^{x-y}\mathrm{I}(x < y < \infty) \end{aligned}$$

One may identify these distributions to the known ones, respectively.

$$\begin{aligned} X &\sim \mathrm{Exp}(1/3) \\ (Y - X)|X &\sim \mathrm{Exp}(1) \end{aligned}$$

Hence

$$\begin{aligned} \mathbb{E}(Y|X) &= \mathbb{E}(Y - X|X) + \mathbb{E}(X|X) = 1 + X \\ \mathrm{Var}(Y|X) &= \mathrm{Var}(Y - X|X) = 1 \end{aligned}$$

As a result, $\mathrm{Var}[\mathbb{E}(Y|X)] = \mathrm{Var}(X) = 1/9$ and $\mathbb{E}\left[\mathrm{Var}(Y|X)\right] = \mathbb{E}(1) = 1$. Furthermore, since $\mathbb{E}(Y|X) = 1 + X$,

$$\mathrm{Var}[X + Y - \mathbb{E}(Y|X)] = \mathrm{Var}(Y - 1) = \mathrm{Var}(Y) = \frac{1}{9} + 1 = \frac{10}{9}$$

by appealing to the law of total variation.

2.6 이재용 (2020)

Suppose $X \perp\!\!\!\perp Y$ and

$$F_X(x) = \begin{cases} \frac{(x+\theta)^n}{(2\theta)^n}, & |x| \leq \theta \\ 1, & x > \theta \\ 0, & x < -\theta \end{cases} \qquad F_Y(y) = \begin{cases} \frac{1-(-y+\theta)^n}{(2\theta)^n}, & |y| \leq \theta \\ 1, & y > \theta \\ 0, & y < -\theta \end{cases}$$

Compute $\mathbb{E}(X - Y)$.

2.6.1 Answer

Very easy. Left to the tutees. The independence condition is not necessary. Find $\mathbb{E}(X + \theta)$ and $\mathbb{E}(-Y + \theta)$, respectively. Then $\mathbb{E}(X - Y) = \mathbb{E}(X + \theta) + \mathbb{E}(-Y + \theta) - 2\theta$.

2.7 이재용 (2020)

5명의 투표자가 있는 선거구가 있고, 국회의원 후보 A와 B가 있다고 하자. 5명의 투표자 중 M 명이 A후보에게 투표했는데, M이 따르는 확률분포는

$$\mathbb{P}(M = m) = \frac{1}{6} \qquad m = 0, 1, 2, 3, 4, 5$$

라고 하자 투표가 끝난 후, 개표를 시작하여 2명의 투표를 개표하였더니, 이 중 1명은 A에게, 다른 1명은 B에게 투표하였다. 이때 5명의 투표자 중 3명 이상이 A 후보에게 투표했을 확률은 얼마인가?

2.7.1 Answer

첫 2명의 투표 중 A가 득표한 수를 X라고 나타내자.

- (Prior) $M \sim \mathrm{Unif}\{0, 1, 2, 3, 4, 5\}$
- (Model) $X|M \sim \mathrm{HyperGeo}(5, M, 2)$
- (Posterior) $M|X \sim ?$

특히 본 문제는 $M|X = 1$ 이라는 posterior distribution에 대해 묻는 것이 된다. 분포 $\mathrm{HyperGeo}(5, M, 2)$ 는 다음과 같이 주어진다. Indicator function에 들어갈 X 의 support에 주의한다.

$$\mathbb{P}(X = x|M = m) = \frac{\binom{2}{x}\binom{3}{m-x}}{\binom{5}{m}} \mathrm{I}_{\{\max(0,m-3),\cdots,\min(2,m)\}}(x)$$

Bayes' Theorem에 의하여 각 $x = 0, 1, 2$ 에 대하여

$$\begin{aligned}\mathbb{P}(M = m|X = x) &\propto \mathbb{P}(M = m)\mathbb{P}(X = x|M = m) \\ &\propto \mathbb{P}(X = x|M = m) \\ &\propto \frac{\binom{2}{x}\binom{3}{m-x}}{\binom{5}{m}} \mathrm{I}_{\{x,x+1,x+2,x+3\}}(m) \qquad \text{(with respect to } m\text{)}\end{aligned}$$

이 성립한다. 여기서는 M 의 support에 주의한다. 특별히 $x = 1$ 인 경우에는

$$\mathbb{P}(M = m|X = 1) \propto \frac{\binom{3}{m-1}}{\binom{5}{m}} \mathrm{I}_{\{1,2,3,4\}}(m)$$

이므로 합이 1이 되도록 normalize하여

$$\begin{aligned}\mathbb{P}(M = 1|X = 1) = \mathbb{P}(M = 4|X = 1) = 2/10 \\ \mathbb{P}(M = 2|X = 1) = \mathbb{P}(M = 3|X = 1) = 3/10\end{aligned}$$

을 얻는다. 따라서 $\mathbb{P}(M \geq 3|X = 1) = 5/10 = 1/2$.

2.8 김우철 (2017)

Suppose the mgf of X exists. The k-th moment m_k of X is given by

$$m_k = (-1)^k \sum_{l=1}^{k} \sum_{\substack{j_1 \geq 1 \\ j_1+\cdots+j_l=k}} \cdots \sum_{j_l \geq 1} \binom{k}{j_1, \cdots, j_l} \binom{-2}{l} 2^l$$

for all $k = 1, 2, \cdots$. Find the skewness, kurtosis, and pdf of X.

2.8.1 Answer

Consider the mgf $M(t)$ of X. For some $\epsilon > 0$, the following holds for $t \in (-\epsilon, \epsilon)$.

$$\begin{aligned}
M(t) &= 1 + \sum_{k=1}^{\infty} m_k \frac{t^k}{k!} \\
&= 1 + \sum_{k=1}^{\infty} (-1)^k \sum_{l=1}^{k} \underbrace{\sum_{j_1 \geq 1} \cdots \sum_{j_l \geq 1}}_{j_1+\cdots+j_l=k} \binom{k}{j_1, \cdots, j_l} \binom{-2}{l} 2^l \frac{t^k}{k!} \\
&= 1 + \sum_{k=1}^{\infty} \sum_{l=1}^{k} \underbrace{\sum_{j_1 \geq 1} \cdots \sum_{j_l \geq 1}}_{j_1+\cdots+j_l=k} \binom{-2}{l} 2^l \frac{(-t)^k}{j_1! \cdots j_l!} \\
&= 1 + \sum_{k=1}^{\infty} \sum_{l=1}^{k} \binom{-2}{l} 2^l \underbrace{\sum_{j_1 \geq 1} \cdots \sum_{j_l \geq 1}}_{j_1+\cdots+j_l=k} \frac{(-t)^{j_1}}{j_1!} \cdots \frac{(-t)^{j_l}}{j_l!} \\
&= 1 + \sum_{l=1}^{\infty} \binom{-2}{l} 2^l \sum_{j_1 \geq 1} \cdots \sum_{j_l \geq 1} \frac{(-t)^{j_1}}{j_1!} \cdots \frac{(-t)^{j_l}}{j_l!} \\
&= 1 + \sum_{l=1}^{\infty} \binom{-2}{l} 2^l \left(\sum_{j=1}^{\infty} \frac{(-t)^j}{j!} \right)^l \\
&= 1 + \sum_{l=1}^{\infty} \binom{-2}{l} 2^l \left(e^{-t} - 1\right)^l \\
&= \left(1 + 2\left(e^{-t} - 1\right)\right)^{-2} \\
&= \left(\frac{\frac{1}{2} e^t}{1 - \frac{1}{2} e^t} \right)^2,
\end{aligned}$$

which is identical to the mgf of $\mathrm{Negbin}(2, 1/2)$. Now take logarithm to get the cgf.

$$\begin{aligned}
C(t) = \log M(t) &= -2 \log\left(1 + 2(e^{-t} - 1)\right) \\
&= -2 \log(1 - \underbrace{\left(2t - t^2 + \frac{t^3}{3} - \frac{t^4}{12} + O(t^5)\right)}_{Ⓐ}) \qquad (e^{-t} - 1 = -t + t^2/2 - t^3/6 + t^4/24 + \cdots) \\
&= 2Ⓐ + Ⓐ^2 + \frac{2}{3}Ⓐ^3 + \frac{1}{2}Ⓐ^4 + O(t^5) \qquad (-\log(1-x) = x + x^2/2 + x^3/3 + x^4/4 + \cdots) \\
&= \left(4t - 2t^2 + \frac{2}{3}t^3 - \frac{1}{6}t^4 + O(t^5)\right) + \left(4t^2 - 4t^3 + \frac{7}{3}t^4 + O(t^5)\right) + \left(\frac{16}{3}t^3 - 8t^4 + O(t^5)\right) + \left(8t^4 + O(t^5)\right) \\
&= 4t + 2t^2 + 2t^3 + \frac{13}{6}t^4 + O(t^5) \\
&= 4t + \frac{4t^2}{2!} + \frac{12t^3}{3!} + \frac{52t^4}{4!} + O(t^5),
\end{aligned}$$

which implies that $c_1 = 4, c_2 = 4, c_3 = 12, c_4 = 52$. Double check here: https://www.wolframalpha.com/input?i2d=true&i=series+-2log%5C%2840%291%2B2%5C%2840%29exp%5C%2840%29-x%5C%

2841%29-1%5C%2841%29%5C%2841%29. Hence, by Exercises 1.19 and 1.20,

$$\text{skew}(X) = \frac{c_3}{c_2^{3/2}} = \frac{3}{2},$$
$$\text{kurt}(X) = \frac{c_4}{c_2^2} = \frac{13}{4}.$$

From the definition of Negative-binomial distribution, the pdf f of X is given by

$$\begin{aligned} f(x) &= \binom{x-1}{2-1}\left(\frac{1}{2}\right)^2\left(1-\frac{1}{2}\right)^{x-2} \mathrm{I}_{\{2,3,4,\cdots\}}(x) \\ &= (x-1)2^{-x}\mathrm{I}_{\{2,3,4,\cdots\}}(x) \end{aligned}$$

- NOTE: 제가 시험장에 있었다면, 시간 상 pdf는 못 구하고, skewness, kurtosis까지는 구했을 것 같습니다.

1 Lebesgue-Stieltjes Integral and Law of the Unconscious Statistician

1.1 Motivation

Recall that a random variable of mixed type does not assume its pdf. Then how do we define its expectation? For instance, consider the following cdf.

$$F_X(x) = \begin{cases} 1 - \frac{1}{2}e^{-3x}, & x \geq 2 \\ \frac{1}{2} - \frac{1}{2}e^{-3x}, & 0 \leq x < 2 \\ 0, & x < 0 \end{cases}$$

Note that F_X is NOT differentiable since it is not even continuous at $x = 2$. In fact, the cdf F_X of mixed type random variable X is a 50-50 mixture of the following two distributions.

$$F_{X_d}(x) = \mathrm{I}_{[2,\infty)}(x) \qquad (X_d \sim \text{The \textbf{Dirac Delta} distribution concentrated at } x = 2)$$
$$F_{X_c}(x) = (1 - e^{-3x})\mathrm{I}_{[0,\infty)}(x) \qquad (X_c \sim \text{The \textbf{Standard Exponential} distribution of mean 1/3})$$

Henceforth, it is very natural to define the expectation by $\mathbb{E}(X) = \frac{1}{2}\mathbb{E}(X_d) + \frac{1}{2}\mathbb{E}(X_c) = \frac{7}{6}$. Again, how do we define its expectation without assuming the existence of pdf?

1.2 Simple Random Variables

One of the **simplest** random variable is an **indicator** random variable. Given a probability space $(S, \mathcal{F}, \mathbb{P})$ and an event $E \in \mathcal{F}$, an indicator random variable I_E is a function $S \to \mathbb{R}$ defined by

$$\mathrm{I}_E(s) = \begin{cases} 1, & s \in E \\ 0, & s \notin E \end{cases} \qquad s \in S$$

We say **an event** E **occurred** if $\mathrm{I}_E(s) = 1$ and **did not occur** otherwise. Now we are going to handle a **finite linear combination** of these indicator random variables.

Definition 1 (Simple Random Variable)**.** *Given a probability space $(S, \mathcal{F}, \mathbb{P})$, a **nonnegative** random variable $X : S \to \mathbb{R}$ is said to be **simple** if $X = \sum_{j=1}^n a_j \mathrm{I}_{E_j}$ for some $n \in \mathbb{N}$, $a_j \geq 0$, and events $E_j \in \mathcal{F}$. That is,*

$$X(s) = \sum_{j=1}^n a_j \mathrm{I}_{E_j}(s), \qquad s \in S$$

For example, if $|S| < \infty$ and $\mathcal{F} = \mathcal{P}(S)$, then every random variable X is simple.

Definition 2 (Lebesgue-Stieltjes Integral of a Simple Random Variable)**.** *For a **simple** random variable $X = \sum_{j=1}^n a_j \mathrm{I}_{E_j}$, we define the **expectation** of X with respect to $\mathbb{P}$ by*

$$\mathbb{E}_{\mathbb{P}}(X) = \sum_{j=1}^n a_j \mathbb{P}(E_j).$$

*Mathematicians usually denote the left hand side by $\int_S X\,d\mathbb{P}$. It is called the **Lebesgue-Stieltjes integral** of X with respect to $\mathbb{P}$.*

For example, if $E \in \mathcal{F}$ is an event, then $\mathbb{E}_{\mathbb{P}}(\mathrm{I}_E) = \mathbb{P}(E)$. There are possibly many representations for a simple random variable, however the Lebesgue-Stieltjes integral is well-defined by the axioms of $\mathbb{P}$. That is, for $E_1, \cdots, E_n, F_1, \cdots, F_m \in \mathcal{F}$,

$$\text{If } \sum_{j=1}^{n} a_j \mathrm{I}_{E_j} = \sum_{i=1}^{m} b_i \mathrm{I}_{F_i}, \qquad \text{then } \sum_{j=1}^{n} a_j \mathbb{P}(E_j) = \sum_{i=1}^{m} b_i \mathbb{P}(F_i).$$

1.3 Nonnegative Random Variables

Proposition 1. *Every **nonnegative** random variable X can be represented by a **pointwise limit** of a **monotone increasing** sequence of **simple** random variables. That is, there exists a sequence $\{X_i\}_{i=1}^{\infty}$ such that*

$$\begin{aligned} &\lim_{i\to\infty} X_i(s) = X(s) && \forall s \in S && \text{(pointwise convergence)} \\ &X_1(s) \le X_2(s) \le \cdots && \forall s \in S && \text{(monotone increasing)} \\ &X_i \text{ is simple} && \forall i \in \mathbb{N} && \text{(simple)} \end{aligned}$$

Proof. We give a brief sketch here. See **명제 10.2.4** in **해석개론 (김, 김, 계)** for a rigor. Recall that $X^{-1}(a,b] = \{s \in S : a < X(s) \le b\} = (a < X \le b)$ is an event. Hence define

$$\begin{aligned} X_1 &= \mathrm{I}_{X^{-1}(1,\infty)}, \\ X_2 &= \frac{1}{2}\mathrm{I}_{X^{-1}(\frac{1}{2},\frac{2}{2}]} + \frac{2}{2}\mathrm{I}_{X^{-1}(\frac{2}{2},\frac{3}{2}]} + \frac{3}{2}\mathrm{I}_{X^{-1}(\frac{3}{2},2]} + 2\mathrm{I}_{X^{-1}(2,\infty)}, \\ X_3 &= \sum_{k=1}^{12} \frac{k-1}{4}\mathrm{I}_{X^{-1}(\frac{k-1}{4},\frac{k}{4}]} + 3\mathrm{I}_{X^{-1}(3,\infty)}, \end{aligned}$$

and so on. Then $X_1 \le X_2 \le \cdots$ are the desired simple random variables. □

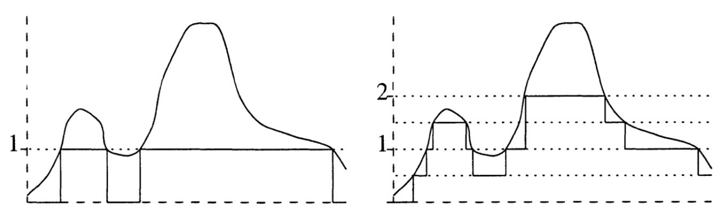


Figure 1: Visualization of X_1 and X_2

We have already defined the Lebesgue-Stieltjes integral of simple random variables. Hence one can apply the definition to $X_1 \le X_2 \le \cdots$ presented above. Verify that $\mathbb{P}(X^{-1}(a,b]) = \mathbb{P}(a < X \le b) = F_X(b) - F_X(a)$.

Following the notion of cdf F_X, the Lebesgue-Stieltjes integrals of $X_1 \le X_2 \le \cdots$ are given as

$$\mathbb{E}_{\mathbb{P}}(X_1) = (1 - F_X(1)),$$
$$\mathbb{E}_{\mathbb{P}}(X_2) = \frac{1}{2}\left(F_X(\tfrac{2}{2}) - F_X(\tfrac{1}{2})\right) + \frac{2}{2}\left(F_X(\tfrac{3}{2}) - F_X(\tfrac{2}{2})\right) + \frac{3}{2}\left(F_X(2) - F_X(\tfrac{3}{2})\right) + 2(1 - F_X(2)),$$
$$\mathbb{E}_{\mathbb{P}}(X_3) = \sum_{k=1}^{12} \frac{k-1}{4}\left(F_X(\tfrac{k-1}{4}) - F_X(\tfrac{k}{4})\right) + 3(1 - F_X(3)),$$

and so on. It seems very similar to the **Riemann-Stieltjes** integral presented in **Section 5.5** of **해석개론 (김, 김, 계)**. Now we are ready to define Lebesgue-Stieltjes integral of a nonnegative random variable.

Definition 3 (Lebesgue-Stieltjes Integral of a Nonnegative Random Variable)**.** *Suppose X is a **nonnegative** random variable. Let $\{X_i\}_{i=1}^{\infty}$ be a monotone increasing sequence of simple random variables that converges to X pointwise (as in the **Proposition 1**). Then the expectation (i.e, Lebesgue-Stieltjes integral) of X with respect to $\mathbb{P}$ is defined by*

$$\mathbb{E}_{\mathbb{P}}(X) = \lim_{i\to\infty} \mathbb{E}_{\mathbb{P}}(X_i)$$

Note that this integral may not be finite.

The **Definitions 2 and 3** coincide for a simple random variable.

Theorem 1 (Monotone Convergence Theorem)**.** *The above Lebesgue-Stieltjes integral is well-defined.*

Proof. See **정리 10.3.1** in **해석개론 (김, 김, 계)**. □

This theorem asserts that the **Definition 3** does NOT depend on the choice of $\{X_i\}_{i=1}^{\infty}$.

1.4 General Random Variables

Definition 4 (Lebesgue-Stieltjes Integral of a General Random Variable)**.** *Suppose $X : S \to \mathbb{R}$ is a random variable. Then the expectation (i.e, Lebesgue-Stieltjes integral) of X with respect to $\mathbb{P}$ is defined by*

$$\mathbb{E}_{\mathbb{P}}(X) = \mathbb{E}_{\mathbb{P}}(X^+) - \mathbb{E}_{\mathbb{P}}(X^-)$$

if the two terms on the right hand side are both finite*.*

Proposition 2. *$\mathbb{E}_{\mathbb{P}}(X)$ is defined if and only if $\mathbb{E}_{\mathbb{P}}(|X|) < \infty$.*

Proof. $|X| = X^+ + X^-$. □

Proposition 3. *Fix a real number x. Then, $\mathbb{E}_{\mathbb{P}}(X\mathrm{I}_{X^{-1}\{x\}}) = x\mathbb{P}(X^{-1}\{x\}) = x\mathbb{P}(X = x)$.*

Proof. We assume $x \ge 0$ first. Consider a constant sequence of simple random variables $X_i = x\mathrm{I}_{X^{-1}\{x\}}$ that converges to $X\mathrm{I}_{X^{-1}\{x\}}$ pointwise. $\mathbb{E}_{\mathbb{P}}(X_i) = x\mathbb{P}(X^{-1}\{x\})$ for all $i = 1, 2, \cdots$. A similar argument is valid for the case $x < 0$. □

Definition 5 (Absolute Continuity of a Random Variable)**.** *A random variable $X : S \to \mathbb{R}$ is said to be* ***absolutely continuous on an open interval*** (a, b) *if the cdf F_X of X is* ***absolutely continuous on the open interval****, i.e, there exists a nonnegative function f_X such that*

$$F_X(x) - F_X(a) = \int_a^x f_X(t)\,dt$$

holds for all $x \in (a, b)$.
A random variable is said to be ***absolutely continuous*** *if it is* ***absolutely continuous*** *on the entire line $\mathbb{R}$. In this case, f_X is called the* ***pdf*** *of X. Mathematicians says f_X is the* ***Radon-Nikodym derivative*** *of F_X.*

Here are some remarks regarding absolute continuity.

- Note that **absolute continuity** is a bit weaker than **differentiablity** and a bit stronger than **continuity**.
- Every **continuous, piecewise differentiable** function is **absolutely continuous**.

Now we present an analogy to **정리 5.5.5** in **해석개론 (김, 김, 계)**.

Theorem 2. *Suppose a random variable $X : S \to \mathbb{R}$ is absolutely continuous on an open interval (a, b). Then,*

$$\mathbb{E}_{\mathbb{P}}(X \mathrm{I}_{X^{-1}(a,b)}) = \int_a^b x f_X(x)\,dx.$$

In particular, if X is absolutely continuous (on the entire line $\mathbb{R}$), then

$$\mathbb{E}_{\mathbb{P}}(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx.$$

Proof. Beyond the scope of undergraduate analysis. □

1.5 Back to the Beginning

Now we are able to rigorously compute the expectation $\mathbb{E}(X)$ where its cdf F_X is given by

$$F_X(x) = \begin{cases} 1 - \frac{1}{2}e^{-3x}, & x \geq 2 \\ \frac{1}{2} - \frac{1}{2}e^{-3x}, & 0 \leq x < 2 \\ 0, & x < 0 \end{cases}$$

(even if its pdf f_X does not exist.) Verify that F_X is absolutely continuous on $(-\infty, 2)$ with its derivative $\frac{3}{2}e^{-3x}\mathrm{I}_{(0,2)}(x)$ and also on $(2, \infty)$ with its derivative $\frac{3}{2}e^{-3x}\mathrm{I}_{(2,\infty)}(x)$. Hence

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}}(X) &= \mathbb{E}_{\mathbb{P}}(X\mathrm{I}_{X^{-1}(-\infty,2)}) + \mathbb{E}_{\mathbb{P}}(X\mathrm{I}_{X^{-1}\{2\}}) + \mathbb{E}_{\mathbb{P}}(X\mathrm{I}_{X^{-1}(2,\infty)}) \\
&= \int_{-\infty}^{2} x\frac{3}{2}e^{-3x}\mathrm{I}_{(0,2)}(x)dx + 2\mathbb{P}(X = 2) + \int_2^\infty x\frac{3}{2}e^{-3x}\mathrm{I}_{(2,\infty)}(x)dx \\
&= 2\mathbb{P}(X = 2) + \frac{3}{2}\int_0^\infty xe^{-3x}dx \\
&= 1 + \frac{1}{6} = \frac{7}{6}
\end{aligned}$$

1.6 Law of the Unconscious Statistician (Very Optional)

Lemma 1 (Lebesgue-Stieltjes Probability on the Real Line)**.** *Given a probability space $(S, \mathcal{F}, \mathbb{P})$, suppose $X : S \to \mathbb{R}$ is a random variable. Define $\mathcal{B}$ and $\mathbb{P}_X$ by*

$$\mathcal{B} = \left\{B \in \mathcal{P}(\mathbb{R}) : X^{-1}(B) \in \mathcal{F}\right\}$$
$$\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B)). \qquad (B \in \mathcal{B})$$

Then, $\mathbb{P}_X((a, b]) = F_X(b) - F_X(a)$ for all $a < b$ and $(\mathbb{R}, \mathcal{B}, \mathbb{P}_X)$ is indeed a probability space.

Proof. (i) $\emptyset \in \mathcal{B}$ since $X^{-1}(\emptyset) = \emptyset \in \mathcal{F}$.
(ii) If $B \in \mathcal{B}$, then $\mathbb{R} \setminus B \in \mathcal{B}$ since $X^{-1}(\mathbb{R} \setminus B) = X^{-1}(\mathbb{R}) \setminus X^{-1}(B) = S \setminus X^{-1}(B) \in \mathcal{F}$.
(iii) If $\{B_j\}_{j=1}^{\infty} \subseteq \mathcal{B}$, then $\bigcup_{j=1}^{\infty} B_j \in \mathcal{B}$ since

$$X^{-1}\left(\bigcup_{j=1}^{\infty} B_j\right) = \bigcup_{j=1}^{\infty} X^{-1}(B_j) \in \mathcal{F}.$$

(iv) $\mathbb{P}_X(\emptyset) = \mathbb{P}(X^{-1}(\emptyset)) = \mathbb{P}(\emptyset) = 0$.
(v) $\mathbb{P}_X(\mathbb{R}) = \mathbb{P}(X^{-1}(\mathbb{R})) = \mathbb{P}(S) = 1$.
(vi) If $\{B_j\}_{j=1}^{\infty} \subseteq \mathcal{B}$ are disjoint events, then

$$\mathbb{P}_X\left(\bigcup_{j=1}^{\infty} B_j\right) = \mathbb{P}\left(X^{-1}\left(\bigcup_{j=1}^{\infty} B_j\right)\right) = \mathbb{P}\left(\bigcup_{j=1}^{\infty} X^{-1}(B_j)\right) = \sum_{j=1}^{\infty} \mathbb{P}(X^{-1}(B_j)) = \sum_{j=1}^{\infty} \mathbb{P}_X(B_j).$$

From (i)-(vi), $(\mathbb{R}, \mathcal{B}, \mathbb{P}_X)$ is a probability space. In particular, $\mathbb{P}_X((a, b]) = \mathbb{P}(X^{-1}(a, b]) = F_X(b) - F_X(a)$. □

Theorem 3 (Law of the Unconscious Statistician)**.** *Let $u : \mathbb{R} \to \mathbb{R}$ be a continuous real function. Then $u \circ X : S \to \mathbb{R}$ and $u : \mathbb{R} \to \mathbb{R}$ are random variables defined in the probability spaces $(S, \mathcal{F}, \mathbb{P})$ and $(\mathbb{R}, \mathcal{B}, \mathbb{P}_X)$, respectively. In particular, one has $\mathbb{E}_{\mathbb{P}}(u \circ X) = \mathbb{E}_{\mathbb{P}_X}(u)$, or equivalently,*

$$\mathbb{E}_{\mathbb{P}}(u \circ X) = \int_{-\infty}^{\infty} u \, d\mathbb{P}_X.$$

Proof. For simplicity, we assume u is nonnegative here. For $B \in \mathcal{B}$, it is obvious that $\mathrm{I}_B \circ X = \mathrm{I}_{X^{-1}(B)}$ and hence

$$\mathbb{E}_{\mathbb{P}}(\mathrm{I}_B \circ X) = \mathbb{E}_{\mathbb{P}}(\mathrm{I}_{X^{-1}(B)}) = \mathbb{P}(X^{-1}(B)) = \mathbb{P}_X(B) = \mathbb{E}_{\mathbb{P}_X}(\mathrm{I}_B).$$

Let $\{u_i\}_{i=1}^{\infty}$ be a monotone increasing sequence of simple functions that converges to u pointwise. Then $\{u_i \circ X\}_{i=1}^{\infty}$ is monotone increasing and converges to $u \circ X$ pointwise. Therefore,

$$\mathbb{E}_{\mathbb{P}}(u \circ X) = \lim_{i \to \infty} \mathbb{E}_{\mathbb{P}}(u_i \circ X) = \lim_{i \to \infty} \mathbb{E}_{\mathbb{P}_X}(u_i) = \mathbb{E}_{\mathbb{P}_X}(u).$$

The continuity assumption of u is necessary to ensure that $(u \circ X)^{-1}((-\infty, x]) \in \mathcal{F}$ for each $x \in \mathbb{R}$. □

As a final remark, statisticians write $\mathbb{E}(u(X)) = \mathbb{E}_{\mathbb{P}}(u \circ X)$ if no confusion can arise. (e.g. $\mathbb{E}(X^2 + \log X)$)

pdf를 작성할 때 support를 반드시 명시해야 한다. 가령 이항분포의 pdf로 올바른 표현은 $\binom{n}{x}p^x(1-p)^{n-x}\mathrm{I}_{\{0,1,\cdots,n\}}(x)$ 이다.

name	notation	support	probability density function	moment generating function
이항분포 Binomial	$\mathrm{B}(n,p)$ $0 \le p \le 1$	$\{0,1,\cdots,n\}$	$\binom{n}{x}p^x q^{n-x}$ where $q=1-p$	$(pe^t+q)^n$ for $t \in \mathbb{R}$
음이항분포 Neg. Bin.	$\mathrm{Negbin}(r,p)$ $0 \le p \le 1$	$\{r, r+1, \cdots\}$	$\binom{x-1}{r-1}p^r q^{x-r}$ where $q=1-p$	$\left(\frac{pe^t}{1-qe^t}\right)^r$ for $t < -\log q$
포아송분포 Poisson	$\mathrm{Poisson}(\lambda)$ $\lambda \ge 0$	$\{0,1,\cdots\}$	$\frac{e^{-\lambda}\lambda^x}{x!}$	$e^{\lambda(e^t-1)}$ for $t \in \mathbb{R}$
다항분포 Multinomial	$\mathrm{Multi}(n,(p_1,\cdots,p_k)^\top)$ $\sum_{j=1}^k p_j = 1, p_j \ge 0$	①	$\frac{n!}{x_1!\cdots x_k!}p_1^{x_1}\cdots p_k^{x_k}$	$\left(\sum_{j=1}^k p_j e^{t_j}\right)^n$ for $t_j \in \mathbb{R}$
감마분포 Gamma	$\mathrm{Gamma}(\alpha,\beta)$ $\alpha,\beta>0$	$(0,\infty)$	$\frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-x/\beta}$	$(1-\beta t)^{-\alpha}$ for $t<\frac{1}{\beta}$
베타분포 Beta	$\mathrm{Beta}(\alpha,\beta)$ $\alpha,\beta>0$	$(0,1)$	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$	$\mathbb{E}(X^k)=\frac{\Gamma(\alpha+\beta)\Gamma(\alpha+k)}{\Gamma(\alpha)\Gamma(\alpha+\beta+k)}$ mgf exists for $t \in \mathbb{R}$
베타이항분포 Beta Bin.	$\mathrm{Betabin}(n,\alpha,\beta)$ $\alpha,\beta>0$	$\{0,1,\cdots,n\}$	$\binom{n}{x}\frac{\Gamma(\alpha+\beta)\Gamma(\alpha+x)\Gamma(\beta+n-x)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha+\beta+n)}$	mgf exists for $t \in \mathbb{R}$
역감마분포 Inv. Gamma	$\mathrm{invGamma}(\alpha,\beta)$ $\alpha,\beta>0$	$(0,\infty)$	$\frac{1}{\Gamma(\alpha)\beta^\alpha}x^{-\alpha-1}e^{-\frac{1}{\beta x}}$	mgf does not exist
로지스틱분포 Logistic	$\mathrm{L}(\mu,\sigma)$ $\mu\in\mathbb{R},\sigma>0$	$(-\infty,\infty)$	$\frac{e^{-z}}{\sigma(1+e^{-z})^2}$ where $z=\frac{x-\mu}{\sigma}$	$e^{\mu t}\Gamma(1-\sigma t)\Gamma(1+\sigma t)$ for $-\frac{1}{\sigma}<t<\frac{1}{\sigma}$
정규분포 Normal	$\mathrm{N}(\mu,\sigma^2)$ $\mu\in\mathbb{R},\sigma>0$	$(-\infty,\infty)$	$\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$e^{\mu t+\frac{1}{2}\sigma^2t^2}$ for $t \in \mathbb{R}$
로그정규분포 Log Norm.	$\mathrm{Lognormal}(\mu,\sigma^2)$ $\mu\in\mathbb{R},\sigma>0$	$(0,\infty)$	$\frac{1}{x\sqrt{2\pi}\sigma}\exp\left(-\frac{(\log x-\mu)^2}{2\sigma^2}\right)$	$\mathbb{E}(X^k)=e^{k\mu+\frac{1}{2}k^2\sigma^2}$ HOWEVER mgf does not exist
$t-$분포 Student's t	t_ν $\nu>0$	$(-\infty,\infty)$	$\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$	mgf does not exist
웨이불분포 Weibull	$\mathrm{Weibull}(\alpha,\beta)$ $\alpha,\beta>0$	$(0,\infty)$	$\frac{\alpha}{\beta^\alpha}x^{\alpha-1}e^{-(x/\beta)^\alpha}$	$\mathbb{E}(X^k)=\beta^k\Gamma(1+\frac{k}{\alpha})$ mgf exists if $\alpha \ge 1$

질문. 베르누이(Bernoulli), 기하(Geometric), 지수(Exponential), 카이제곱(χ^2), 코시(Cauchy)분포를 이 표에서 찾을 수 있겠는가?

Notes

Parameters

$$f(x;\mu,\sigma) = \frac{1}{\sigma} f_0\left(\frac{x-\mu}{\sigma}\right)$$

꼴이면, μ, σ를 각각 location, scale parameter라고 부른다. 꼭 평균, 표준편차일 필요는 없다. θ^{-1}가 scale parameter인 경우 보통 θ를 rate parameter라고 부른다. 나머지 경우 일반적으로 shape parameter라고 부른다.

Gamma Integral

$\alpha > 0$에 대하여 감마함수는 다음과 같이 정의된다.

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

먼저 부분적분을 통해 $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$를 보일 수 있다:

$$\begin{aligned}\Gamma(\alpha+1) &= \int_0^\infty x^{\alpha} e^{-x} dx \\ &= \left[x^\alpha e^{-x}\right]_\infty^0 + \int_0^\infty \alpha x^{\alpha-1} e^{-x} dx = \alpha\Gamma(\alpha)\end{aligned}$$

$\Gamma(1) = 1$이므로 $n = 0, 1, \cdots$ 일 때 $\Gamma(n+1) = n!$ 임을 알 수 있다.

Beta Integral

$\alpha, \beta > 0$에 대하여

$$\begin{aligned}\Gamma(\alpha)\Gamma(\beta) &= \int_0^\infty x^{\alpha-1} e^{-x} dx \int_0^\infty y^{\beta-1} e^{-y} dy \\ &= \int_0^\infty \int_0^\infty x^{\alpha-1} y^{\beta-1} e^{-x-y} dx dy\end{aligned}$$

여기서 $x = zw, y = z(1-w)$ 치환하면 $\dfrac{\partial(x,y)}{\partial(z,w)} = \begin{vmatrix} w & z \\ 1-w & -z \end{vmatrix} = -z$이므로

$$\begin{aligned}\Gamma(\alpha)\Gamma(\beta) &= \int_0^1 \int_0^\infty (zw)^{\alpha-1} (z(1-w))^{\beta-1} e^{-z} z dz dw \\ &= \int_0^1 w^{\alpha-1}(1-w)^{\beta-1} dw \underbrace{\int_0^\infty z^{\alpha+\beta-1} e^{-z} dz}_{=\Gamma(\alpha+\beta)}\end{aligned}$$

따라서 $\alpha, \beta > 0$에 대하여 베타함수를 $\mathrm{B}(\alpha,\beta) = \dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ 라고 정의하면

$$\mathrm{B}(\alpha,\beta) = \int_0^1 w^{\alpha-1}(1-w)^{\beta-1} dw$$

가 성립한다. *질문.* $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ 임을 보일 수 있겠는가? *Hint.* $\theta \in [0, \frac{\pi}{2}]$에 대하여 $w = \sin^2\theta$.

Derivation

$X \sim \mathrm{B}(n,p)$ 이면 $t \in \mathbb{R}$ 에 대하여

$$\mathbb{E}(e^{tX}) = \sum_{x=0}^{n} e^{tx}\binom{n}{x}p^x q^{n-x} = \sum_{x=0}^{n}\binom{n}{x}(pe^t)^x q^{n-x} = (pe^t+q)^n$$

$X \sim \mathrm{Negbin}(r,p)$ 이면 $qe^t < 1$ 일 때

$$\mathbb{E}(e^{tX}) = \sum_{x=r}^{\infty} e^{tx}\binom{x-1}{r-1}p^r q^{x-r} = \sum_{x=r}^{\infty}\binom{x-1}{r-1}(pe^t)^r(qe^t)^{x-r} = \left(\frac{pe^t}{1-qe^t}\right)^r \underbrace{\sum_{x=r}^{\infty}\binom{x-1}{r-1}(1-qe^t)^r(qe^t)^{x-r}}_{=1}$$

$X \sim \mathrm{Poisson}(\lambda)$ 이면 $t \in \mathbb{R}$ 에 대하여

$$\mathbb{E}(e^{tX}) = \sum_{x=0}^{\infty} e^{tx}\frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{(\lambda e^t)^x}{x!} = e^{-\lambda+\lambda e^t}$$

$X \sim \mathrm{Multi}(n,(p_1,\cdots,p_k)^\top)$ 이면 $t_1,\cdots,t_k \in \mathbb{R}$ 에 대하여

$$\begin{aligned}\mathbb{E}(e^{t_1X_1+\cdots+t_kX_k}) &= \sum_{x_1+\cdots+x_k=n} e^{t_1x_1+\cdots+t_kx_k}\frac{n!}{x_1!\cdots x_k!}p_1^{x_1}\cdots p_k^{x_k} \\ &= \sum_{x_1+\cdots+x_k=n}\frac{n!}{x_1!\cdots x_k!}(p_1e^{t_1})^{x_1}\cdots(p_ke^{t_k})^{x_k} \\ &= \left(p_1e^{t_1}+\cdots+p_ke^{t_k}\right)^n\end{aligned}$$

$X \sim \mathrm{Gamma}(\alpha,\beta)$ 이면 $\beta t < 1$ 일 때

$$\begin{aligned}\mathbb{E}(e^{tX}) &= \int_0^\infty e^{tx}\frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-\frac{1}{\beta}x}dx \\ &= \frac{1}{(1-\beta t)^\alpha}\underbrace{\int_0^\infty \frac{1}{\Gamma(\alpha)}\left(\frac{1}{\beta}-t\right)^\alpha x^{\alpha-1}e^{-(\frac{1}{\beta}-t)x}dx}_{=1}\end{aligned}$$

$X \sim \mathrm{Beta}(\alpha,\beta)$ 이면 $k=1,2,\cdots$ 에 대하여

$$\begin{aligned}\mathbb{E}(X^k) &= \int_0^1 x^k\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}dx \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\left(\frac{\Gamma(\alpha+\beta+k)}{\Gamma(\alpha+k)\Gamma(\beta)}\right)^{-1}\underbrace{\int_0^1\frac{\Gamma(\alpha+\beta+k)}{\Gamma(\alpha+k)\Gamma(\beta)}x^{\alpha+k-1}(1-x)^{\beta-1}dx}_{=1}\end{aligned}$$

와 같이 $k-$th moment를 얻을 수 있을 뿐만 아니라, $t \in \mathbb{R}$ 에 대하여

$$\mathbb{E}(e^{tX}) = \int_0^1 e^{tx}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}dx$$

인데 $x \in [0,1]$ 에서 $e^{tx} \le e^{|t|}$ 이므로 $\mathbb{E}(e^{tX}) \le e^{|t|}$ 이다. 따라서 mgf가 모든 $t \in \mathbb{R}$ 에 대하여 존재하고, 정리 1.5.2에 의거하여

$$\mathbb{E}(e^{tX}) = \sum_{k=0}^{\infty}\frac{\mathbb{E}(X^k)}{k!}t^k = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)}\sum_{k=0}^{\infty}\frac{t^k}{k!}\frac{\Gamma(\alpha+k)}{\Gamma(\alpha+\beta+k)}$$

$p \sim \text{Beta}(\alpha, \beta)$ 이고 $X|p \sim \text{B}(n, p)$ 이면 $x \in \{0, \cdots, n\}$ 에 대하여

$$\begin{aligned}\mathbb{P}(X = x) &= \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}\binom{n}{x}p^x(1-p)^{n-x}dp \\ &= \binom{n}{x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\left(\frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+x)\Gamma(\beta+n-x)}\right)^{-1}\underbrace{\int_0^1 \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+x)\Gamma(\beta+n-x)}p^{\alpha+x-1}(1-p)^{\beta+n-x-1}dp}_{=1}\end{aligned}$$

이고

$$\mathbb{E}(X) = \mathbb{E}[\mathbb{E}(X|p)] = \mathbb{E}[np] = n \cdot \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+1)}{\Gamma(\alpha)\Gamma(\alpha+\beta+1)} = \frac{n\alpha}{\alpha+\beta}$$

일반적으로 $t \in \mathbb{R}$ 에 대하여

$$\begin{aligned}\mathbb{E}(e^{tX}) = \mathbb{E}[\mathbb{E}(e^{tX}|p)] &= \mathbb{E}\left[\left(1+p(e^t-1)\right)^n\right] \\ &= \mathbb{E}\left[\sum_{k=0}^n \binom{n}{k}p^k(e^t-1)^k\right] \\ &= \sum_{k=0}^n \binom{n}{k}\mathbb{E}\left[p^k\right](e^t-1)^k \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)}\sum_{k=0}^n \binom{n}{k}\frac{\Gamma(\alpha+k)}{\Gamma(\alpha+\beta+k)}(e^t-1)^k\end{aligned}$$

$X \sim \text{invGamma}(\alpha, \beta)$ 이면 $t > 0$ 에 대하여

$$\mathbb{E}(e^{tX}) = \int_0^\infty e^{tx}\frac{1}{\Gamma(\alpha)\beta^\alpha}x^{-\alpha-1}e^{-\frac{1}{\beta x}}dx$$

인데 $x \geq M \implies tx - (\alpha+1)\log x - \frac{1}{\beta x} > \frac{t}{2}x$ 를 만족하는 $M > 0$ 이 존재하고

$$\int_M^\infty e^{\frac{t}{2}x}dx = \infty$$

이므로 mgf는 존재하지 않는다. *Note.* 역감마분포에서 $k-$th moment의 존재성은 α 와 k 의 대소와 관련되어 있다. $\alpha > 0$ 이기만 하면 분포가 잘 정의되지만, $k-$th moment를 가지려면 $\alpha > k$ 여야 한다. 이 경우,

$$\begin{aligned}\mathbb{E}(X^k) &= \int_0^\infty x^k\frac{1}{\Gamma(\alpha)\beta^\alpha}x^{-\alpha-1}e^{-\frac{1}{\beta x}}dx \\ &= \frac{\Gamma(\alpha-k)}{\Gamma(\alpha)\beta^k}\underbrace{\int_0^\infty \frac{1}{\Gamma(\alpha-k)\beta^{\alpha-k}}x^{-(\alpha-k)-1}e^{-\frac{1}{\beta x}}dx}_{=1}\end{aligned}$$

$X \sim \text{L}(\mu, \sigma)$ 이면 $|\sigma t| < 1$ 일 때 $z = \frac{x-\mu}{\sigma}$ 에 대하여 $dz = \frac{1}{\sigma}dx$ 이고 $w = \frac{1}{1+e^{-z}}$ 에 대하여 $dw = \frac{e^{-z}}{(1+e^{-z})^2}dz$ 이므로

$$\begin{aligned}\mathbb{E}(e^{tX}) = \int_{-\infty}^\infty e^{tx}\frac{e^{-z}}{\sigma(1+e^{-z})^2}dx &= \int_{-\infty}^\infty e^{t(\mu+\sigma z)}\frac{e^{-z}}{(1+e^{-z})^2}dz \\ &= e^{\mu t}\int_0^1\left(\frac{w}{1-w}\right)^{\sigma t}dw \\ &= e^{\mu t}\Gamma(1-\sigma t)\Gamma(1+\sigma t)\underbrace{\int_0^1 \frac{\Gamma(2)}{\Gamma(1+\sigma t)\Gamma(1-\sigma t)}w^{1+\sigma t-1}(1-w)^{1-\sigma t-1}dw}_{=1}\end{aligned}$$

$X \sim \mathrm{N}(\mu, \sigma^2)$ 이면 $t \in \mathbb{R}$ 일 때 $z = \frac{x-\mu}{\sigma}$ 에 대하여 $dz = \frac{1}{\sigma}dx$ 이므로

$$\begin{aligned}\mathbb{E}(e^{tX}) &= \int_{-\infty}^{\infty} e^{tx}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx = \int_{-\infty}^{\infty} e^{t(\mu+\sigma z)}\frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz \\ &= e^{\mu t+\frac{1}{2}\sigma^2t^2}\underbrace{\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-(z-\sigma t)^2/2}dz}_{=1}\end{aligned}$$

$X \sim \mathrm{Lognormal}(\mu, \sigma^2)$ 이면 $k = 0, 1, \cdots$ 에 대하여

$$\begin{aligned}\mathbb{E}(X^k) &= \int_0^{\infty} x^k\frac{1}{x\sqrt{2\pi}\sigma}\exp\left(-\frac{(\log x-\mu)^2}{2\sigma^2}\right)dx \\ &= \int_{-\infty}^{\infty} e^{ky}\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)dy && (y = \log x) \\ &= e^{k\mu+\frac{1}{2}k^2\sigma^2} && \text{(정규분포의 mgf 유도과정을 다시 살펴보자)}\end{aligned}$$

그럼에도 불구하고 임의의 $t > 0$ 에 대하여 $\mathbb{E}(e^{tX}) = \infty$ 임을 보인 적이 있다.
$X \sim t_\nu$ 면 $t > 0$ 에 대하여

$$\mathbb{E}(e^{tX}) = \int_{-\infty}^{\infty} e^{tx}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}dx$$

인데 역감마함수와 같은 논증으로 mgf는 존재하지 않는다. ($x \to \infty$ 일 때 피적분함수가 발산한다.) *Note.* 역감마분포와 마찬가지로 $k-$th moment의 존재성은 ν 와 k 의 대소와 관련되어 있다. $\nu > 0$ 이기만 하면 분포가 잘 정의되지만, $k-$th moment를 가지려면 $\nu > k$ 여야 한다. 이 경우, pdf가 even function이므로 k 가 odd일 때 $\mathbb{E}(X^k) = 0$ 이고 k 가 even일 때 $x \geq 0$ 에 대하여

$$z = \frac{x^2}{\nu+x^2} \qquad x = \sqrt{\frac{\nu z}{1-z}} \qquad dx = \frac{1}{2}\sqrt{\frac{\nu}{z(1-z)^3}}dz$$

로 치환하면

$$\begin{aligned}\mathbb{E}(X^k) &= \int_{-\infty}^{\infty} x^k\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}dx \\ &= \frac{2\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\int_0^{\infty} x^k\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}dx && \text{(pdf and } k \text{ are even)} \\ &= \frac{2\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\int_0^1\left(\frac{\nu z}{1-z}\right)^{\frac{k}{2}}(1-z)^{\frac{\nu+1}{2}}\frac{1}{2}\sqrt{\frac{\nu}{z(1-z)^3}}dz && \text{(substitute } x \text{ by } z\text{)} \\ &= \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\nu^{\frac{k+1}{2}}\int_0^1 z^{\frac{k+1}{2}-1}(1-z)^{\frac{\nu-k}{2}-1}dz \\ &= \frac{1}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\nu^{\frac{k+1}{2}}\Gamma\left(\frac{k+1}{2}\right)\Gamma\left(\frac{\nu-k}{2}\right) && \text{(Beta Integral)} \\ &= \frac{\Gamma\left(\frac{1}{2}+m\right)\Gamma\left(\frac{\nu}{2}-m\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{\nu}{2}\right)}\nu^m && \text{(Let } m = \tfrac{k}{2} \in \mathbb{Z}\text{)}\end{aligned}$$

다시 한 번 강조하지만, 이 모든 논의는 $\nu > k$ 일 때 가능한 것이다.
$X \sim \mathrm{Weibull}(\alpha, \beta)$ 이면 $(\alpha, \beta > 0)$

$$z = \left(\frac{x}{\beta}\right)^{\alpha} \qquad dz = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1}dx$$

로 치환하여

$$\mathbb{E}(X^k) = \int_0^\infty x^k \frac{\alpha}{\beta^\alpha} x^{\alpha-1} e^{-(x/\beta)^\alpha} dx = \beta^k \int_0^\infty z^{k/\alpha} e^{-z} dz = \beta^k \Gamma\left(1 + \frac{k}{\alpha}\right)$$

를 얻는다.
이상의 논의는 모든 $\alpha, \beta > 0$에 대하여 성립했으나, 웨이블분포의 mgf가 존재하기 위해서는 $\alpha \geq 1$이어야 함이 알려져 있다.
$\alpha > 1$인 경우에는 $t \in \mathbb{R}$에 대하여

$$\mathbb{E}(e^{tX}) = \int_0^\infty e^{tx} \frac{\alpha}{\beta^\alpha} x^{\alpha-1} e^{-(x/\beta)^\alpha} dx = \int_0^\infty \exp\left(\beta t z^{1/\alpha} - z\right) dz$$

인데 $z \geq M \implies \beta t z^{1/\alpha} \leq \frac{z}{2}$가 성립하는 $M > 0$이 존재하고

$$\int_M^\infty \exp\left(-\frac{z}{2}\right) < \infty$$

이므로 mgf가 모든 $t \in \mathbb{R}$에 대하여 존재한다. 이제는 정리 1.5.2에 의거하여

$$\mathbb{E}(e^{tX}) = \sum_{k=0}^\infty \frac{\mathbb{E}(X^k)}{k!} t^k = \sum_{k=0}^\infty \frac{(\beta t)^k}{k!} \Gamma\left(1 + \frac{k}{\alpha}\right)$$

라고 적을 수 있다. 정확하게 $\alpha = 1$인 경우에는 $\beta t < 1$인 t에 대하여 mgf가 존재할 것이다. 그리고 이 경우는 지수분포에 해당한다. (다음 절을 참조하라.)

Related Distributions

$\text{Bernoulli}(p) \overset{d}{=} \text{B}(1, p)$	$p^x(1-p)^{1-x} \text{I}_{\{0,1\}}(x)$	(베르누이 Bernoulli 분포)
$\text{Geo}(p) \overset{d}{=} \text{Negbin}(1, p)$	$p(1-p)^{x-1} \text{I}_{\{1,2,\cdots\}}(x)$	(기하 Geometric 분포)
$\text{Exp}(\beta) \overset{d}{=} \text{Gamma}(1, \beta)$	$\frac{1}{\beta} e^{-x/\beta} \text{I}_{(0,\infty)}(x)$	(지수 Exponential 분포)
$\chi^2_\nu \overset{d}{=} \text{Gamma}\left(\frac{\nu}{2}, 2\right)$	$\frac{1}{\Gamma\left(\frac{\nu}{2}\right) 2^{\nu/2}} x^{\frac{\nu}{2}-1} e^{-x/2} \text{I}_{(0,\infty)}(x)$	(카이제곱 χ^2 분포)
$\text{Cauchy}(0, 1) \overset{d}{=} t_1$	$\frac{1}{\pi(1+x^2)} \text{I}_{(-\infty,\infty)}(x)$	(코시 Cauchy 분포)
$\text{Unif}(0, 1) \overset{d}{=} \text{Beta}(1, 1)$	$\text{I}_{(0,1)}(x)$	(균등 Uniform 분포)

Generalized Gamma Distribution

name	notation	support	probability density function	moment generating function
일반화된 감마분포 Generalized Gamma	$\text{GG}(d, p, \beta)$ $d, p, \beta > 0$	$(0, \infty)$	$\frac{p}{\Gamma(d/p)\beta^d} x^{d-1} e^{-(x/\beta)^p}$	$\mathbb{E}(X^k)$ mgf exists if $p \geq 1$

Note. $p = 1$이면 감마분포가 되고, $d = p$이면 웨이블분포가 되고, $d = 1, p = 2$이면 반정규 Half Normal 분포(정규분포를 따르는 확률변수의 절대값을 생각)가 된다.
$X \sim \text{GG}(d, p, \beta)$이면 $(d, p, \beta > 0)$

$$z = \left(\frac{x}{\beta}\right)^p \qquad dz = \frac{p}{\beta}\left(\frac{x}{\beta}\right)^{p-1} dx$$

로 치환하여 $k = 0, 1, 2, \cdots$ 에 대하여

$$\begin{aligned}\mathbb{E}(X^k) &= \int_0^\infty x^k \frac{p}{\Gamma(d/p)\,\beta^d} x^{d-1} e^{-(x/\beta)^p}\,dx \\ &= \frac{\beta^k}{\Gamma(d/p)} \int_0^\infty \left(\frac{x}{\beta}\right)^{d+k-p} \cdot e^{-(x/\beta)^p} \cdot \frac{p}{\beta}\left(\frac{x}{\beta}\right)^{p-1} dx \\ &= \frac{\beta^k}{\Gamma(d/p)} \int_0^\infty z^{\frac{d+k-p}{p}} e^{-z} dz \\ &= \beta^k \frac{\Gamma\left(\frac{d+k}{p}\right)}{\Gamma\left(\frac{d}{p}\right)}\end{aligned}$$

더 나아가 $p \geq 1$ 이라면 mgf가 존재한다. 정확히 $p = 1$ 이라면 앞서 다룬 감마분포에 해당하게 된다. $p > 1$ 이라면 $t \in \mathbb{R}$ 에 대하여

$$\begin{aligned}\mathbb{E}(e^{tX}) &= \int_0^\infty e^{tx} \frac{p}{\Gamma(d/p)\,\beta^d} x^{d-1} e^{-(x/\beta)^p}\,dx \\ &= \frac{1}{\Gamma(d/p)} \int_0^\infty e^{tx} \cdot \left(\frac{x}{\beta}\right)^{d-p} \cdot e^{-(x/\beta)^p} \cdot \frac{p}{\beta}\left(\frac{x}{\beta}\right)^{p-1} dx \\ &= \frac{1}{\Gamma(d/p)} \int_0^\infty \exp\left(\beta t z^{1/p} + \frac{d-p}{p}\log z - z\right) dz\end{aligned}$$

$M > 0$ 이 존재하여 $z > M$ 이면 $\exp(\cdot)$ 안의 항이 $-z/2$ 보다 작게 된다. 그리고

$$\int_M^\infty \exp\left(-\frac{z}{2}\right) dz < \infty$$

이므로 mgf가 모든 $t \in \mathbb{R}$ 에 대하여 존재한다. 이제 정리 1.5.2를 적용할 수 있다.

CDF and Sampling Theory

$\mathbb{P}(X \leq x) = F(x)$ 이고 $U \sim \text{unif}(0,1)$ 이라고 할 때 X 와 $F^{-1}(U)$ 는 정확히 같은 분포가 됨을 지난 번 `2.pdf`에서 밝혔었다. F 의 inverse가 존재하면 그대로 사용하면 되고, 존재하지 않는다면 다음의 generalized version을 사용하는 것이다.

$$F^{-1}(u) = \inf\{y \in \mathbb{R} : F(y) \geq u\} \qquad 0 < u < 1$$

이것은 항상 잘 정의되는 것을 역시 밝혔었다.
$X \sim \text{Exp}(\beta)$ 면 pdf와 cdf가 각각 $f(x) = \frac{1}{\beta} e^{-x/\beta} \mathrm{I}_{(0,\infty)}(x)$ 와

$$F(x) = \begin{cases} 1 - e^{-x/\beta} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

으로 주어지므로 $X \overset{d}{=} -\beta \log(1-U)$ 가 성립한다.
응용. CDF를 통한 sampling은 로지스틱분포에서도 유효하다.
$X \sim \text{Gamma}(\alpha, \theta), Y \sim \text{Gamma}(\beta, \theta)$ 이고 $X \perp Y$ 이면 X, Y 의 joint pdf는

$$f(x,y) = \frac{1}{\Gamma(\alpha)\Gamma(\beta)\theta^{\alpha+\beta}} x^{\alpha-1} y^{\beta-1} e^{-(x+y)/\theta} \mathrm{I}_{(0,\infty)}(x) \mathrm{I}_{(0,\infty)}(y)$$

로 주어지므로 $x = zw, y = z(1-w)$ 로 치환하여 $W = \frac{X}{X+Y} \sim \text{Beta}(\alpha, \beta)$ 임을 확인할 수 있다.

1 Drills and Skills: Random Vectors and Change of Variables

1.1 Recap: Differential and Regularity

Definition 1 (Differential)**.** *Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a (real) multivariable differentiable function. If one writes F by*

$$F(x_1, \cdots, x_n) = (F_1(x_1, \cdots, x_n), \cdots F_m(x_1, \cdots, x_n)),$$

*then for given $p \in \mathbb{R}^n$, the **differential** dF_p of F at p is an $\mathbb{R}$-linear map $\mathbb{R}^n \to \mathbb{R}^m$ represented by an $m \times n$ matrix,*

$$dF_p = \begin{bmatrix} \frac{\partial F_1}{\partial x_1}(p) & \cdots & \frac{\partial F_1}{\partial x_n}(p) \\ \vdots & \ddots & \vdots \\ \frac{\partial F_m}{\partial x_1}(p) & \cdots & \frac{\partial F_m}{\partial x_n}(p) \end{bmatrix},$$

with respect to the standard coordinate of Euclidean spaces.

- Note: There are a number of equivalent notations for the differential.

$$dF_p = d_pF = DF_p = D_pF = \frac{\partial F}{\partial x}(p) = \frac{\partial(F_1, \cdots, F_m)}{\partial(x_1, \cdots, x_n)}(p) = J_F(p) = \nabla F(p) = F'(p) = \dot{F}(p) = \cdots$$

Definition 2 (Regularity)**.** *Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a (real) multivariable differentiable function. A point p in the domain, i.e, $\mathbb{R}^n$ is said to be a **regular point** of F if dF_p is surjective, that is,*

$$\operatorname{rank} dF_p = \dim \operatorname{im} dF_p = m.$$

*A value c in the codomain, i.e, $\mathbb{R}^m$ is said to be a **regular value** of F if $F^{-1}(c) = \emptyset$ or every point in $F^{-1}(c)$ is **regular**. A point that is not **regular** is called **critical**. A value that is not **regular** is called **critical**.*

- Example: Consider $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = 2x^3 - 3x^2$. Then, 0 and 1 are the only critical points; 0 and -1 are the only critical values.
- Example: Consider $f : \mathbb{R}^2 \to \mathbb{R}$ given by $f(x, y) = x^2 - y^2$. Then, $(0, 0)$ is the only critical point; 0 is the only critical value.

1.2 Recap: Inverse Function Theorem

Now we focus on the special case $m = n$. In this case, as described in the Linear Algebra class, given $p \in \mathbb{R}^n$, the followings are equivalent:

- dF_p is surjective, i.e, p is a regular point by definition.
- dF_p is of full rank, namely, n.
- dF_p is invertible.
- $\det dF_p \neq 0$.
- dF_p is an $\mathbb{R}$-vector space isomorphism.

In fact, the Inverse Function Theorem says more than this.

Theorem 1 (Inverse Function Theorem). *Suppose $F : \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable. If dF_p is invertible for some $p \in \mathbb{R}^n$, then F is a local C^1-diffeomorphism at p. That is, there exists an open neighborhood U of p such that $F|_U : U \to F(U)$ has its inverse $F^{-1} : F(U) \to U$ which is continuously differentiable. Moreover, for all $c \in F(U)$, the inverse F^{-1} satisfies*

$$(dF^{-1})_c = (dF_{F^{-1}(c)})^{-1}.$$

- The theorem writes in a more familiar way for the case $n = 1$:

$$(f^{-1})'(c) = \frac{1}{f'(f^{-1}(c))}.$$

1.3 Random Vectors and Change of Variables

A **random vector** is defined in a **canonical** way. To elaborate on this, for each $j = 1, \cdots, n$, consider a function $\pi_j : \mathbb{R}^n \to \mathbb{R}$ defined by

$$\pi_j(x_1, \cdots, x_n) = x_j$$

Such π_j is called the **canonical projection**, or equivalently, **canonical surjection**.

Definition 3 (Random Vector). *Given a probability space $(S, \mathcal{F}, \mathbb{P})$, an n-**dimensional random vector (-차원 확률 벡터)** or n-**variate random variable (-변량 확률 변수)** is a function $X : S \to \mathbb{R}^n$ such that $\pi_j \circ X$ is a random variable for all $j = 1, \cdots, n$.*

- Note: In analogy to the case $n = 1$, an n-dimensional random variable is called absolutely continuous if

$$\mathbb{P}(X \le (x_1, \cdots, x_n)) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_X(t_1, \cdots, t_n)\, dt_n \cdots dt_1$$

  for some $f_X : \mathbb{R}^n \to \mathbb{R}_{\ge 0}$, which is called the **pdf** of X.

Theorem 2 (Change of Variables). *Suppose $X : S \to \mathbb{R}^n$ is an absolutely continuous n-dimensional random vector endowed with a pdf f_X. If an n-dimensional (real) continuously differentiable function $u : \mathbb{R}^n \to \mathbb{R}^n$ is defined **almost everywhere** and assumes **almost every** regular point, i.e,*

$$\mathbb{P}(\|u(X)\| < \infty) = 1, \qquad \mathbb{P}(\det du_X \ne 0) = 1,$$

then $Y := u \circ X$ is an absolutely continuous n-dimensional random vector endowed with a pdf f_Y given by

$$f_Y(y) = \sum_{x \in u^{-1}(y)} \frac{f_X(x)}{|\det du_x|},$$

*which is defined for all regular values $y \in \mathbb{R}^n$ of u. This pdf is well-defined since Y is regular **almost surely**.*

Proof. See Theorem 2.47 in (Folland, 1999). (It is beyond the scope of undergraduate calculus and analysis.) Or equivalently, see **정리 4.1.2 (다대일 변환을 통한 확률변수의 치환법) in the textbook (수리통계학)**. □

- Note: One may memorize the formula in an intuitive way: $f_Y(y)|dy| = f_X(x)|dx|$.
- Note: If y is a regular value of u, then $\det du_x \neq 0$ for all $x \in u^{-1}(y)$ by definition. Hence, the fraction on right hand side of the theorem makes sense.
- Remark: The assumption that X is regular almost surely is essential. Consider the following example.

$$Y = u(X), X \sim \mathrm{N}(0,1), u(x) = e^{-1/x}\mathrm{I}_{(0,\infty)}(x)$$

Then, u is an element in $C^\infty(\mathbb{R})$ (the space of real smooth functions), i.e, has derivatives of all orders at all points $x \in \mathbb{R}$. However, $Y = u(X)$ may and does not admit a pdf since only positive points x are regular and $\mathbb{P}(X > 0) \neq 1$. Can you identify the cdf of Y instead?

2 Exercises: One-Dimensional

THE BASIS OF YOUR NEW KNOWLEDGE SHOULD BE YOUR PREVIOUS KNOWLEDGE.

2.1

Suppose the pdf of a random variable X is given by

$$f_X(x) = \frac{1}{2}\mathrm{I}_{(-1,1)}(x) \qquad \text{(called the Uniform distribution supported on } (-1,1))$$

Find the pdf of $Y = X^2$. Can you identify the distribution to a known one?

2.1.1 ANSWER

Let $u : (-1,1) \to \mathbb{R}$ be defined by $y = u(x) = x^2$. Observe that $\mathbb{P}(Y = u(X) \in (0,1)) = 1$ and $u^{-1}(y) = \{-\sqrt{y}, \sqrt{y}\}$ for all $y \in (0,1)$. Hence one has

$$f_Y(y) = f_X(-\sqrt{y})\left|-\frac{d}{dy}\sqrt{y}\right| + f_X(\sqrt{y})\left|\frac{d}{dy}\sqrt{y}\right| = \frac{1}{2\sqrt{y}}\mathrm{I}_{(0,1)}(y),$$

which is the pdf of $\mathrm{Beta}(1/2, 1)$.

2.2

Suppose the pdf of a random variable X is given by

$$f_X(x) = e^{-x}\mathrm{I}_{(0,\infty)}(x) \qquad \text{(called the standard Exponential distribution)}$$

Find the pdf of $Y = \frac{1}{(\log X)^2}$.

2.2.1 ANSWER

Define $u : (0, \infty) \setminus \{1\} \to \mathbb{R}$ by $y = u(x) = 1/(\log x)^2$. After checking some regularity conditions for $y > 0$, one has

$$\begin{aligned} f_Y(y) &= f_X\left(e^{-1/\sqrt{y}}\right)\left|\frac{d}{dy}e^{-1/\sqrt{y}}\right| + f_X\left(e^{1/\sqrt{y}}\right)\left|\frac{d}{dy}e^{1/\sqrt{y}}\right| \\ &= \frac{1}{2y\sqrt{y}}\left(e^{-e^{-1/\sqrt{y}}}e^{-1/\sqrt{y}} + e^{-e^{1/\sqrt{y}}}e^{1/\sqrt{y}}\right)\mathrm{I}_{(0,\infty)}(y). \end{aligned}$$

2.3

Suppose the pdf of a random variable X is given by

$$f_X(x) = \frac{1}{\pi(1+x^2)}\mathrm{I}_{(-\infty,\infty)}(x) \qquad \text{(called the standard Cauchy distribution)}$$

Find the pdf of $Y = \frac{X^2}{1+X^2}$. Can you identify the distribution to a known one?

2.3.1 ANSWER

It is easily verified that $Y \sim \mathrm{Beta}(1/2, 1/2)$.

2.4

Suppose the pdf of a random variable X is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2} \qquad \text{(called the standard Normal distribution)}$$

Find the pdfs of $Y = e^X$ and $Z = X^2$, respectively. Can you identify the distributions to known ones?

2.4.1 ANSWER

$Y \sim$ the standard log-normal distribution. $Z \sim$ the χ^2 distribution with degree of freedom 1.

2.5

Suppose the pdf of a random variable X is f_X. Find the pdf of $Y = \mu + \sigma X$ for given $\mu \in \mathbb{R}, \sigma > 0$.

2.6

Suppose the pdf and cdf of a random variable X are given by f_X and F_X, respectively. Find the pdf of $Y = F_X(X)$. Assume f_X is continuous and does not vanish everyehere, i.e, $f_X > 0$.

2.6.1 ANSWER

$$f_Y(y) = \frac{1}{\sigma}f_X\left(\frac{y-\mu}{\sigma}\right)$$

3 Exercises: Multi-Dimensional

YOU WILL BECOME MUCH STRONGER BY EMBRACING YOUR VULNERABILITIES.

3.1

Suppose the joint pdf of random variables (X, Y) is given by

$$f_{X,Y}(x, y) = 2e^{-x-2y}\mathrm{I}_{(0,\infty)}(x)\mathrm{I}_{(0,\infty)}(y) \qquad \text{(independent Exponential distributions)}$$

Find the pdfs of $Z = \min(X, Y)$ and $W = \max(X, Y)$, respectively. Describe the distributions.

3.1.1 ANSWER

Consider a function $u : (0, \infty)^2 \to \mathbb{R}^2$ that maps (X, Y) to (Z, W). Since $\mathbb{P}(Z < W) = 1$, one has

$$f_{Z,W}(z, w) = f_{X,Y}(z, w) + f_{X,Y}(w, z) = 2\left(e^{-z-2w} + e^{-2z-w}\right)\mathrm{I}(0 < z < w < \infty)$$

Some integrations show us that

$$f_Z(z) = \int_z^\infty 2\left(e^{-z-2w} + e^{-2z-w}\right)\,dw = 3e^{-3z}\mathrm{I}_{(0,\infty)}(z)$$
$$f_W(w) = \int_0^z 2\left(e^{-z-2w} + e^{-2z-w}\right)\,dz = \left(2\left(e^{-2w} - e^{-3w}\right) + e^{-w} - e^{-3w}\right)\mathrm{I}_{(0,\infty)}(w)$$

Note that $Z = \min(X, Y)$ follows the Exponential distribution with the summed rate $= 3$.

3.2

Suppose the joint pdf of random variables (X, Y) is given by

$$f_{X,Y}(x, y) = \frac{1}{12}x^2y^3e^{-x-y}\mathrm{I}_{(0,\infty)}(x)\mathrm{I}_{(0,\infty)}(y) \qquad \text{(independent Gamma distributions with common rates)}$$

Find the pdfs of $Z = X + Y$ and $W = \frac{X}{X+Y}$, respectively. Describe the distributions.

3.2.1 ANSWER

Consider a function $u : (0, \infty)^2 \to \mathbb{R}^2$ that maps (X, Y) to (Z, W). By restricting the codomain of u, its inverse is well-defined by

$$u^{-1}(z, w) = (zw, z(1 - w)) \qquad ((z, w) \in (0, \infty) \times (0, 1))$$

The differential of inverse evaluated at (z, w) is given by

$$(du^{-1})_{(z,w)} = \begin{bmatrix} w & z \\ 1 - w & -z \end{bmatrix},$$

which has the determinant of $-z$. It follows that

$$f_{Z,W}(z,w) = f_X(zw, z(1-w))z = \frac{1}{12}(zw)^2(z(1-w))^3e^{-z}z = \frac{1}{720}z^6e^{-z}\mathrm{I}_{(0,\infty)}(z) \cdot 60w^2(1-w)^3\mathrm{I}_{(0,1)}(w).$$

Hence Z and W are independent. $Z \sim \mathrm{Gamma}(7,1)$ and $W \sim \mathrm{Beta}(3,4)$.

3.3

Let $X_1, \cdots, X_n$ be iid standard Uniform samples. That is,

$$f_{X_1,\cdots,X_n}(x_1,\cdots,x_n) = \mathrm{I}_{(0,1)}(x_1)\cdots\mathrm{I}_{(0,1)}(x_n)$$

Rearrange the samples in a non-decreasing way, say, $X_{(1)} \leq \cdots \leq X_{(n)}$. Find the joint pdf of $X_{(1)} \leq \cdots \leq X_{(n)}$. Find the marginal pdf of $X_{(j)}$ for each j. Describe the distributions.

3.3.1 Answer

$$f_{X_{(1)},\cdots,X_{(n)}}(x_{(1)},\cdots,x_{(n)}) = n!\,\mathrm{I}(0 < x_{(1)} < \cdots < x_{(n)} < 1)$$

and

$$f_{X_{(j)}}(x_{(j)}) = \frac{n!}{(j-1)!(n-j)!}(x_{(j)})^{j-1}(1-x_{(j)})^{n-j}\mathrm{I}_{(0,1)}(x_{(j)})$$

hold. That is, $X_{(j)} \sim \mathrm{Beta}(j, n-j+1)$ for $1 \leq j \leq n$.

3.4

Suppose the joint pdf of random variables (X,Y) is given by

$$f_{X,Y}(x,y) = \frac{1}{30\pi}y^{5/2}e^{-(x^2+y)/2}\mathrm{I}_{(-\infty,\infty)}(x)\mathrm{I}_{(0,\infty)}(y) \qquad \text{(independent Normal and Gamma distributions)}$$

Find the pdfs of X, Y, and $Z = \frac{X}{\sqrt{Y/7}}$, respectively. *Hint: Consider a function u that maps (X,Y) to (X,Z).*

3.4.1 Answer

It is a bit easier to consider a function $u : (X,Y) \mapsto (Z,Y)$, not (X,Z). (My appologies...) Then u admits its inverse defined by

$$u^{-1}(z,y) = (x,y) = \left(z\sqrt{y/7}, y\right).$$

Compute the differential and its determinant at (z,y).

$$(du^{-1})_{(z,y)} = \begin{bmatrix} \sqrt{y/7} & \frac{z}{2\sqrt{7y}} \\ 0 & 1 \end{bmatrix}, \qquad |\det(du^{-1})_{(z,y)}| = \sqrt{y/7}.$$

It follows that

$$\begin{aligned} f_{Z,Y}(z,y) &= \frac{1}{30\pi} y^{5/2} \exp\left(-\frac{(z\sqrt{y/7})^2 + y}{2}\right)\sqrt{\frac{y}{7}} \\ &= \frac{1}{30\pi\sqrt{7}} y^3 \exp\left(-\frac{1+z^2/7}{2}y\right) \mathrm{I}_{(-\infty,\infty)}(z)\mathrm{I}_{(0,\infty)}(y). \end{aligned}$$

Recall that $\int_0^\infty y^3 e^{-\lambda y} dy = \Gamma(4)\lambda^{-4} = 6\lambda^{-4}$. Integrating out y gives the marginal pdf of Z, as desired.

$$\begin{aligned} f_Z(z) &= \int_0^\infty f_{Z,Y}(z,y)\,dy \\ &= \int_0^\infty \frac{1}{30\pi\sqrt{7}} y^3 \exp\left(-\frac{1+z^2/7}{2}y\right)\,dy \\ &= \frac{6}{30\pi\sqrt{7}}\left(\frac{1+z^2/7}{2}\right)^{-4} \\ &= \frac{16}{5\pi\sqrt{7}}\left(1+\frac{z^2}{7}\right)^{-4} \mathrm{I}_{(-\infty,\infty)}(z). \end{aligned}$$

- Note: In general, pdf of the Student's t-distribution with degree of freedom ν is given by

$$f_Z(z) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{z^2}{\nu}\right)^{-\frac{\nu+1}{2}}.$$

For the case $\nu = 7$,

$$\frac{\Gamma(4)}{\sqrt{7\pi}\,\Gamma\left(\frac{7}{2}\right)} = \frac{6}{\sqrt{7\pi}\frac{5}{2}\cdot\frac{3}{2}\cdot\frac{1}{2}\sqrt{\pi}} = \frac{16}{5\pi\sqrt{7}}.$$

References

Folland, G. B. (1999). *Real analysis: modern techniques and their applications* (Vol. 40). John Wiley & Sons.

1 Drills, Drills, and Drills

1.1 김우철 (2015)

Suppose $X \sim \text{Unif}(-2, 3)$. Find the pdf of Y for the following cases, respectively.

(a) $Y = 3 + 2\log\dfrac{2+X}{3-X}$ (b) $Y = 3\left(-\log\dfrac{3-X}{5}\right)^{1/2}$ (c) $Y = X^2$

1.1.1 Answer

$$
\begin{aligned}
f_Y(y) &= \frac{\exp\left(\frac{y-3}{2}\right)}{2\left(1+\exp\left(\frac{y-3}{2}\right)\right)^2}\mathrm{I}_{(-\infty,\infty)}(y),\\
f_Y(y) &= \frac{2}{9}ye^{-y^2/9}\mathrm{I}_{(0,\infty)}(y),\\
f_Y(y) &= \begin{cases} \frac{1}{5\sqrt{y}}, & y \in (0,4) \\ \frac{1}{10\sqrt{y}}, & y \in (4,9) \\ 0, & otherwise \end{cases}
\end{aligned}
$$

1.2 김우철 (2016)

Suppose $X, Y \sim iid\,\text{Geo}(p)$.

(a) Prove that U and V are independent where $U = \min(X, Y)$ and $V = X - Y$.

(b) Find the distribution of $Z = \frac{X}{X+Y}$.

1.2.1 Answer

$$
\begin{aligned}
f_{U,V}(u,v) = \left\{\begin{matrix} f_{X,Y}(u,u), & v = 0 \\ f_{X,Y}(u+v,u), & v > 0 \\ f_{X,Y}(u,u-v), & v < 0 \end{matrix}\right\} &= (1-p)^{2u+|v|-2}p^2\mathrm{I}_{\{1,2,\cdots\}}(u)\mathrm{I}_{\mathbb{Z}}(v) \\
&= \left((2p-p^2)(1-p)^{2u-2}\mathrm{I}_{\{1,2,\cdots\}}(u)\right)\left(\frac{p}{2-p}(1-p)^{|v|}\mathrm{I}_{\mathbb{Z}}(v)\right)
\end{aligned}
$$

For $m, n \in \{1, 2, \cdots\}$ such that $\gcd(m, n) = 1$,

$$
f_Z\left(\frac{m}{m+n}\right) = \sum_{k=1}^{\infty} f_{X,Y}(mk, nk) = \sum_{k=1}^{\infty}(1-p)^{(m+n)k-2}p^2 = \frac{(1-p)^{m+n-2}p^2}{1-(1-p)^{m+n}}.
$$

1.3 김우철 (2016)

Suppose $X \sim \mathrm{Exp}(\lambda)$ and $Y \sim \mathrm{Exp}(\mu)$ are independent. Define Z and W by

$$Z = \min(X, Y) \qquad\qquad W = \begin{cases} 1, & Z = X \\ 0, & Z = Y \end{cases}$$

(a) Find the joint distribution of Z and W.

(b) Prove that Z and W are independent.

1.3.1 SKETCH OF ANSWER

We can NOT apply Change of Variables here. *Hint*: For $z > 0$, verify that

$$\mathbb{P}(Z > z|W = 1) = \mathbb{P}(X > z|Y > X) = \frac{\mathbb{P}(Y > X > z)}{\mathbb{P}(Y > X)} = \frac{\int_z^\infty \int_x^\infty f_{X,Y}(x,y)dydx}{\int_0^\infty \int_x^\infty f_{X,Y}(x,y)dydx}$$
$$\mathbb{P}(Z > z|W = 0) = \mathbb{P}(Y > z|X > Y) = \frac{\mathbb{P}(X > Y > z)}{\mathbb{P}(X > Y)} = \frac{\int_z^\infty \int_y^\infty f_{X,Y}(x,y)dxdy}{\int_0^\infty \int_y^\infty f_{X,Y}(x,y)dxdy}$$

Compare the two quantities.

1.4 김우철 (2015)

Suppose $X_1, X_2 \sim iid\, \mathrm{N}(0,1)$.

(a) Find the joint pdf of $Y_1 = X_1^2 + X_2^2$ and $Y_2 = X_1X_2/Y_1$.

(b) Find the pdf of $Z = X_1/(X_1 + X_2)$.

1.4.1 ANSWER

(a) Consider a map $u : (x_1, x_2) \mapsto \left(x_1^2 + x_2^2, \frac{x_1x_2}{x_1^2+x_2^2}\right)$. Obeserve that if $(y_1, y_2) = u(x_1, x_2)$ for $x_1, x_2 \in \mathbb{R}$ such that $x_1^2 - x_2^2 \neq 0$, then $u^{-1}(y_1, y_2) = \{(x_1, x_2), (x_2, x_1), (-x_1, -x_2), (-x_2, -x_1)\}$. In addition, one has

$$du_{(x_1,x_2)} = \frac{\partial(y_1, y_2)}{\partial(x_1, x_2)} = \begin{bmatrix} 2x_1 & 2x_2 \\ \frac{x_2(x_2^2-x_1^2)}{(x_1^2+x_2^2)^2} & \frac{x_1(x_1^2-x_2^2)}{(x_1^2+x_2^2)^2} \end{bmatrix}$$

and hence

$$|\det du_{(x_1,x_2)}| = \frac{2|x_1^2 - x_2^2|}{x_1^2 + x_2^2},$$

all of which coincide for $(x_1, x_2) \in u^{-1}(y_1, y_2)$. Now the Change of Variables formula asserts that

$$\begin{aligned} f_{Y_1,Y_2}(y_1, y_2) &= \frac{f_{X_1,X_2}(x_1, x_2)}{|\det du_{(x_1,x_2)}|} \times 4 \\ &= \frac{1}{2\pi} \exp\left(-\frac{x_1^2 + x_2^2}{2}\right) \frac{2(x_1^2 + x_2^2)}{|x_1^2 - x_2^2|} \\ &= \left(\frac{1}{2} e^{-y_1/2} \mathrm{I}_{(0,\infty)}(y_1)\right) \left(\frac{2}{\pi\sqrt{1 - 4y_2^2}} \mathrm{I}_{(-1/2,1/2)}(y_2)\right). \end{aligned}$$

(b) Consider a map $v : (x_1, x_2) \mapsto (x_1, \frac{x_1}{x_1+x_2})$.

$$f_Z(z) = \frac{1}{\frac{1}{2}\pi\left((\frac{z-1/2}{1/2})^2 + 1\right)} \mathrm{I}_{\mathbb{R}}(z) \qquad (\sim \text{Cauchy}(1/2, 1/2))$$

1.5 Unknown (2007, 2009)

Suppose X_1, X_2 are jointly distributed by

$$f_{1,2}(x_1, x_2) = \frac{1}{\pi} \mathrm{I}(0 < x_1^2 + x_2^2 < 1).$$

Define $Y_1 = \sqrt{X_1^2 + X_2^2}$ and $Y_2 = X_1/Y_1$.
(a) Find the joint pdf of Y_1 and Y_2.
(b) Find $\mathrm{Cov}(Y_1, Y_2)$.

1.5.1 Answer

Consider a map $u : (x_1, x_2) \mapsto (\sqrt{x_1^2 + x_2^2}, x_1/\sqrt{x_1^2 + x_2^2})$. Then $u^{-1}(y_1, y_2) = \{(y_1 y_2, \pm y_1\sqrt{1 - y_2^2})\}$ for $(y_1, y_2) \in (0, 1) \times (-1, 1)$ and

$$f_{Y_1,Y_2}(y_1, y_2) = \left(2y_1 \mathrm{I}_{(0,1)}(y_1)\right) \left(\frac{1}{\pi\sqrt{1 - y_2^2}} \mathrm{I}_{(-1,1)}(y_2)\right).$$

Since Y_1 and Y_2 are independent, the covariance is zero.

1.6 이재용 (2009, 2020)

X_1, X_2, X_3 are jointly distributed by

$$f_{1,2,3}(x, y, z) = 90e^{-(x+2y+3z)} \mathrm{I}(0 < x < y < z < \infty).$$

Prove or disprove: $X_1, X_2 - X_1, X_3 - X_2$ are mutually independent.
Note: There are at least two techniques you can apply. One is the joint mgf. The other is the change of variables.

1.6.1 Answer Omitted

1.7 김우철 (2016)

Suppose $Z_1, \cdots, Z_K$ are mutually independent and satisfy $Z_i \sim \text{Gamma}(\alpha_i, \beta)$ for each $i = 1, \cdots, K$.
(a) Prove that

$$\left(\frac{Z_1}{\sum_1^K Z_i}, \cdots, \frac{Z_{K-1}}{\sum_1^K Z_i}\right) \sim \text{Dir}\left(\alpha_1, \cdots, \alpha_K\right).$$

(b) Suppose $W_1 \sim \text{Dir}(\omega_1, \cdots, \omega_K), W_2 \sim \text{Dir}(\nu_1, \cdots, \nu_K), V \sim \text{Beta}\left(\sum_1^K \omega_i, \sum_1^K \nu_i\right)$. Assume in addition that W_1, W_2, V are mutually independent. Define Z by

$$Z = VW_1 + (1-V)W_2.$$

Prove that $Z \sim \text{Dir}(\omega_1 + \nu_1, \cdots, \omega_K + \nu_K)$.
(c) Suppose $\mathbf{Y} = (Y_1, \cdots, Y_{K-1}) \sim \text{Dir}(\alpha_1, \cdots, \alpha_K)$. For each $i = 1, \cdots, K-1$, prove that Y_i and

$$\mathbf{Y}_{-i} = \left(\frac{Y_1}{1-Y_i}, \cdots, \frac{Y_{i-1}}{1-Y_i}, \frac{Y_{i+1}}{1-Y_i}, \cdots, \frac{Y_{K-1}}{1-Y_i}\right)$$

are independent.

1.7.1 Answer

(a) See Lecture Note.
(b) Suppose that $\Omega_i \sim \text{Gamma}(\omega_i, 1)$ and $N_i \sim \text{Gamma}(\nu_i, 1)$ for each $i = 1, \cdots, K$ and that they are all mutually independent. Define $\Omega = \sum_1^K \Omega_i$ and $N = \sum_1^K N_i$. Then one has

$$\begin{aligned} W_1 &\overset{d}{=} \left(\frac{\Omega_1}{\Omega}, \cdots, \frac{\Omega_{K-1}}{\Omega}\right) \perp\!\!\!\perp \Omega \sim \text{Gamma}(\sum_1^K \omega_i, 1) \\ W_2 &\overset{d}{=} \left(\frac{N_1}{N}, \cdots, \frac{N_{K-1}}{N}\right) \perp\!\!\!\perp N \sim \text{Gamma}(\sum_1^K \nu_i, 1) \\ V &\overset{d}{=} \frac{\Omega}{\Omega + N} \end{aligned}$$

As a consequence,

$$Z = VW_1 + (1-V)W_2 = \left(\frac{\Omega_1 + N_1}{\Omega + N}, \cdots, \frac{\Omega_{K-1} + N_{K-1}}{\Omega + N}\right) \sim \text{Dir}(\omega_1 + \nu_1, \cdots, \omega_K + \nu_K)$$

since $\Omega_i + N_i \sim \mathrm{Gamma}(\omega_i + \nu_i, 1)$ for each $i = 1, \cdots, K$.
(c) Suppose $A_i \sim \mathrm{Gamma}(\alpha_i, 1)$ are mutually independent for each $i = 1, \cdots, K$ and write

$$Y_j = \frac{A_j}{\sum_{i=1}^K A_i}$$

for $j = 1, \cdots, K-1$. We now prove the statement only for $\mathbf{Y}_{-1}$ without loss of generality.

$$\begin{aligned}\mathbf{Y}_{-1} &= \left(\frac{Y_2}{1-Y_1}, \cdots, \frac{Y_{K-1}}{1-Y_1}\right) \\ &= \left(\frac{A_2}{\sum_{i=2}^K A_i}, \cdots, \frac{A_{K-1}}{\sum_{i=2}^K A_i}\right) \sim \mathrm{Dir}(\alpha_2, \cdots, \alpha_K)\end{aligned}$$

and $\mathbf{Y}_{-1} \perp\!\!\!\perp (A_1, \sum_{i=2}^K A_i)$. Observe that Y_1 is given by a function of A_1 and $\sum_{i=2}^K A_i$. Concretely,

$$Y_1 = \frac{A_1}{\sum_{i=1}^K A_i} = \frac{A_1}{A_1 + \sum_{i=2}^K A_i}$$

holds. It concludes that $\mathbf{Y}_{-1} \perp\!\!\!\perp Y_1$.

0 Preliminaries (Common) - Matrix Series and Exponentiation Map

Let A be any $n \times n$ matrix. We define the matrix exponentiation map $\exp$ by

$$\exp(A) = I_n + \sum_{j=1}^{\infty} \frac{1}{j!} A^j = I_n + A + \frac{1}{2}A^2 + \frac{1}{6}A^3 + \cdots$$

Proposition 1. *Suppose $A = SJS^{-1}$ is the Jordan canonical form of the matrix A. Then,*

$$\exp(A) = S\exp(J)S^{-1}.$$

Corollary 1 (Jacobi's formula)**.**

$$\det \exp(A) = e^{\mathrm{tr} A}.$$

Proposition 2. *Suppose $\gamma(t) = \exp(tA)$ for $t \in \mathbb{R}$. Then one has $\gamma(s+t) = \gamma(s)\gamma(t), \gamma(0) = I_n$, and*

$$\left.\frac{d^j}{dt^j}\right|_{t=0} \gamma(t) = A^j$$

for all $j = 1, 2, \cdots$.

Proposition 3. *If $AB = BA$, then $\exp(AB) = \exp(BA)$. In particular, $\exp(\lambda A) = e^{\lambda} \exp(A)$ for $\lambda \in \mathbb{R}$.*

Proposition 4. *If the spectral radius of A is strictly lesser than 1, i.e, every (possibly complex) eigenvalue of A has a norm lesser than 1, then*

$$I_n + \sum_{j=1}^{\infty} A^j = (I_n - A)^{-1},$$

$$\sum_{j=1}^{\infty} jA^{j-1} = (I_n - A)^{-2}.$$

- 위의 사실들은 수리통계학을 떠나서 굉장히 잘 알려진 상식입니다. 선형대수학 2 등의 기본적인 강좌는 물론이고, 앞으로 여러 분야에서 튜티 여러분이 접할 일이 있을 것입니다.
- 하지만 수리통계학 교과서에서는 **정리 3.4.2**에도 나타나듯이, 의도적으로 행렬 표현을 숨기고 있습니다.
- 따라서 이 페이지는 지금 100% 이해하지 못하더라도 무방합니다. 그러나, 아래에서 서술할 베르누이 과정과 포아송 과정 사이의 analogy는 분명하게 이해해야 할 것입니다.

1 Introduction to Stochastic Processes (기초 확률과정론)

THE BASIS OF **STOCHASTIC PROCESSES** IS THE GAMBLING THEORY. AS A REASON, ONE OF THE MOST INTRINSIC TOPICS IN **MARKOV CHAINS** IS CALLED "GAMBLER'S RUIN."

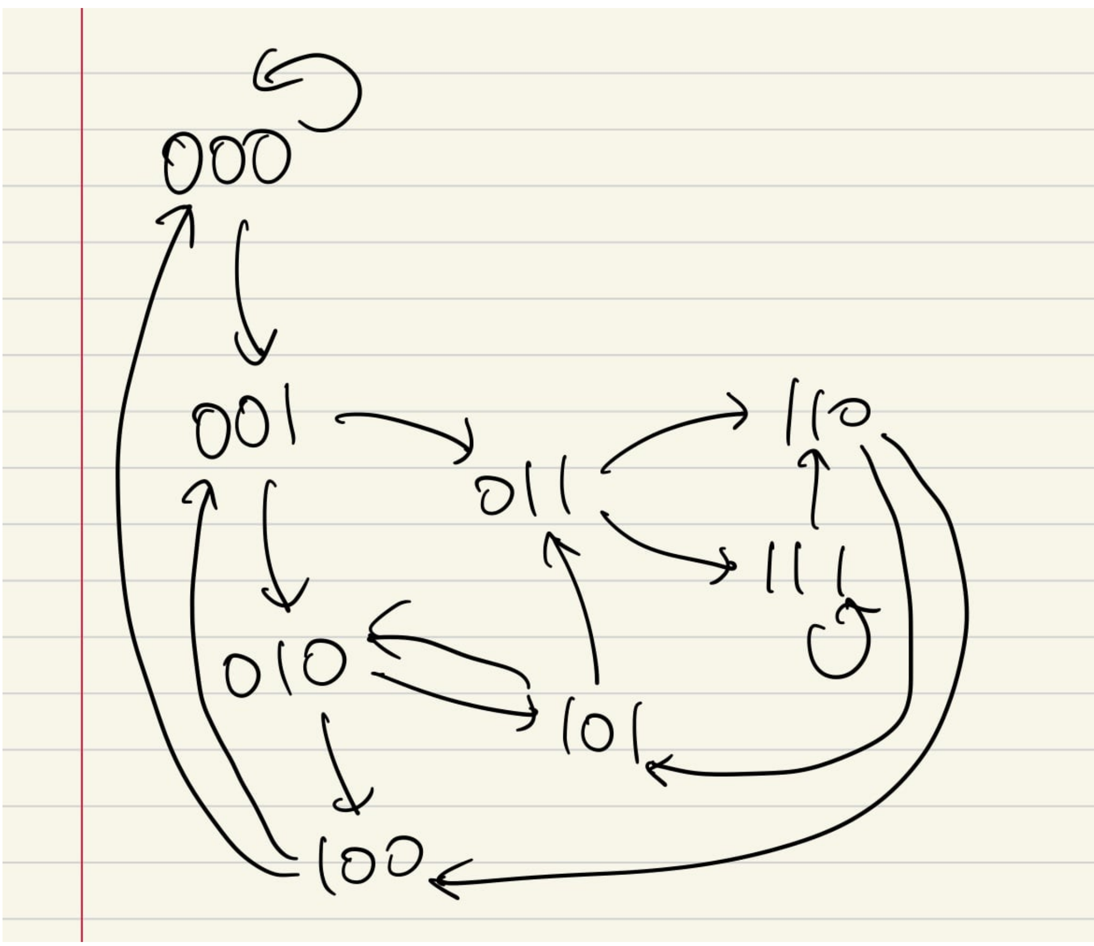
1.1 Markov Chain - Motivation

Pop quiz: Let $(X_1, X_2, \cdots)$ be a sequence of iid Bernoulli random variables with parameter $p = 1/2$. Define

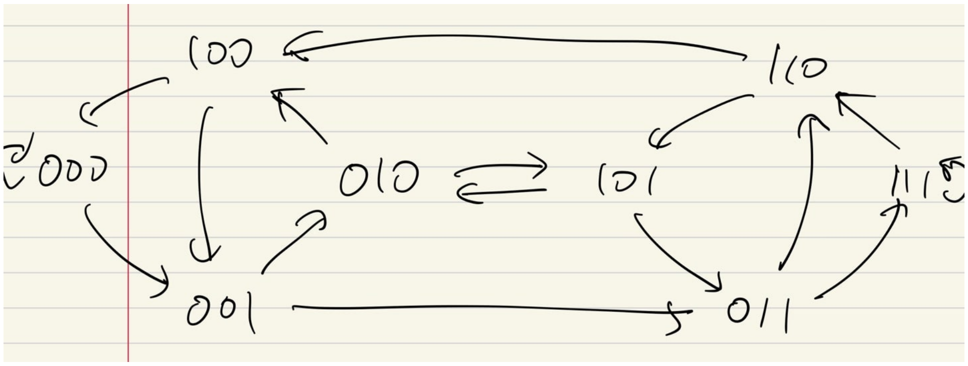
$$W = \min\{t \in \{1, 2, \cdots\} : X_{t-2} = 1, X_{t-1} = 0, X_t = 1\}.$$

Find $\mathbb{E}(W)$. https://math.stackexchange.com/questions/816140/why-is-the-expected-number-coin-tosses-to-get-hth-is-10

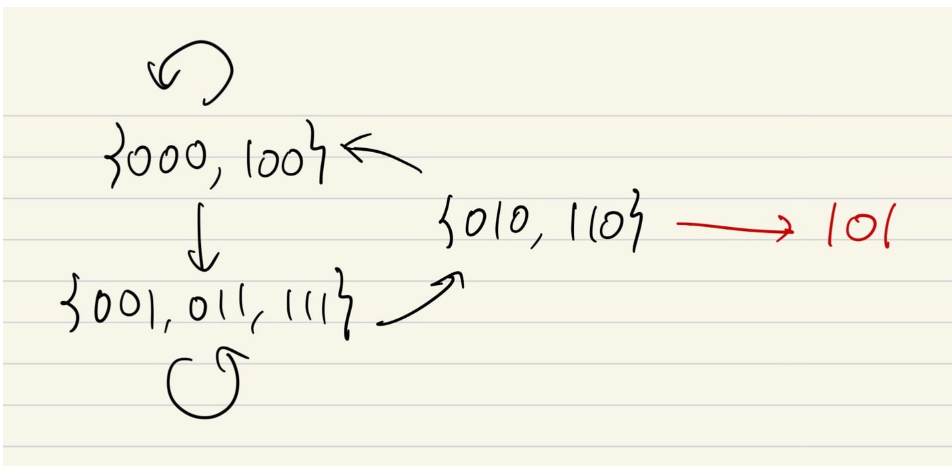
First Attempt. Make a graph that represents the transition probability.



Second Attempt. This looks a bit easier.



Final Attempt.



$$\mathbb{P}(W=w)=\frac{1}{2}\begin{bmatrix}0 & 0 & 1\end{bmatrix}\begin{bmatrix}1/2 & 0 & 1/2\\ 1/2 & 1/2 & 0\\ 0 & 1/2 & 0\end{bmatrix}^{w-1}\begin{bmatrix}1\\0\\0\end{bmatrix}\mathrm{I}_{\{1,2,\cdots\}}(w).$$

The probability decays exponentially w.r.t. w (by the maximal eigenvalue argument). Hence, one can assert that $\mathbb{E}(W)=\sum_w w\mathbb{P}(W=w)<\infty$. Indeedly, by appealing to the **Proposition 4**, one has

$$\begin{aligned}\mathbb{E}(W)&=\frac{1}{2}\begin{bmatrix}0 & 0 & 1\end{bmatrix}\sum_{w=1}^{\infty} w\begin{bmatrix}1/2 & 0 & 1/2\\ 1/2 & 1/2 & 0\\ 0 & 1/2 & 0\end{bmatrix}^{w-1}\begin{bmatrix}1\\0\\0\end{bmatrix}\\ &=\frac{1}{2}\begin{bmatrix}0 & 0 & 1\end{bmatrix}\begin{bmatrix}1/2 & 0 & -1/2\\ -1/2 & 1/2 & 0\\ 0 & -1/2 & 1\end{bmatrix}^{-2}\begin{bmatrix}1\\0\\0\end{bmatrix}=\frac{1}{2}\begin{bmatrix}0 & 0 & 1\end{bmatrix}2^2\begin{bmatrix}2 & 1 & 1\\ 2 & 2 & 1\\ 1 & 1 & 1\end{bmatrix}^{2}\begin{bmatrix}1\\0\\0\end{bmatrix}=10.\end{aligned}$$

1.2 Homogeneous Bernoulli Process

Homogeneous Bernoulli Process (동차 베르누이 과정) of parameter p is a **discrete-time homogeneous counting Markov** process $(N_t)_{t=0}^{\infty} \subseteq \{0, 1, 2, \cdots\}$ with $N_0 = 0$, defined by a **transition probability (전이 확률)** as follows.

$$\mathbb{P}(N_{t+1} = n_{t+1} | N_t = n_t) = \begin{cases} p, & n_{t+1} = n_t + 1 \\ 1-p, & n_{t+1} = n_t \\ 0, & otherwise \end{cases}$$

(Or equivalently, $N_0 = 0$ and $N_t = \sum_{j=1}^{t} X_j$ where $(X_j)_{j=1}^{\infty}$ is a sequence of iid Bernoulli random variables.)

- "Homogeneous" means that p does not depend on time t.
- "Discrete-time" means that $t = 0, 1, 2, \cdots$.
- "Counting" means that $N_t = 0, 1, 2, \cdots$.
- "Markov" means that for all t,

$$\mathbb{P}(N_{t+1} | N_0, N_1, \cdots, N_t) = \mathbb{P}(N_{t+1} | N_t)$$

Let

$$A = \begin{bmatrix} 1-p & 0 & 0 & \cdots \\ p & 1-p & 0 & \cdots \\ 0 & p & 1-p & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \qquad g(t) = \begin{bmatrix} \mathbb{P}(N_t = 0) \\ \mathbb{P}(N_t = 1) \\ \vdots \end{bmatrix}, \qquad g(0) = \begin{bmatrix} 1 \\ 0 \\ \vdots \end{bmatrix}.$$

Then, one has $g(t+1) = Ag(t)$ for all t. Hence,

$$g(t) = A^t g(0) = \begin{bmatrix} (1-p)^t \\ t(1-p)^{t-1}p \\ \frac{t(t-1)}{2}(1-p)^{t-2}p^2 \\ \vdots \end{bmatrix}. \qquad \text{(can be proved via induction on } t\text{)}$$

Related distributions: Write $W_r = \min\{t : N_t \geq r\}$ (Waiting Time)

- Binomial distribution: $N_t \sim \text{Bin}(t, p)$
- Discrete Uniform distribution: $W_1 | N_T = 1 \sim \text{Unif}\{1, 2, \cdots, T\}$
- Hypergeometric distribution: $N_t | N_T = r \sim \text{Hypergeo}(r, T, t)$
- Geometric distribution: $W_1 \sim \text{Geo}(p)$
- Negative binomial distribution: $W_r \sim \text{NegBin}(r, p)$

1.3 Homogeneous Poisson Process

Homogeneous Poisson Process (동차 포아송 과정) of parameter λ is a **continuous-time homogeneous counting Markov** process $(N_t : t \geq 0) \subseteq \{0, 1, 2, \cdots\}$ with $N_0 = 0$, defined by a **transition probability (전이 확률)** as follows.

$$\mathbb{P}(N_{t+h} = n_{t+h} | N_t = n_t) = \begin{cases} \lambda h + o(h), & n_{t+h} = n_t + 1 \\ 1 - \lambda h + o(h), & n_{t+h} = n_t \\ o(h), & otherwise \end{cases}$$

- "Homogeneous" means that λ does not depend on time t.
- "Continuous-time" means that $t \in [0, \infty)$.
- "Counting" means that $N_t = 0, 1, 2, \cdots$.
- "Markov" means that for all $t, h > 0$,

$$\mathbb{P}(N_{t+h} | N_s, s \leq t) = \mathbb{P}(N_{t+h} | N_t)$$

Let

$$A = \begin{bmatrix} -\lambda & 0 & 0 & \cdots \\ \lambda & -\lambda & 0 & \cdots \\ 0 & \lambda & -\lambda & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \qquad g(t) = \begin{bmatrix} \mathbb{P}(N_t = 0) \\ \mathbb{P}(N_t = 1) \\ \vdots \end{bmatrix}, \qquad g(0) = \begin{bmatrix} 1 \\ 0 \\ \vdots \end{bmatrix}.$$

Then, one has $g(t + h) = (I + hA)g(t) + o(h)$ for all $t, h \geq 0$. Hence, $g'(t) = Ag(t)$ and

$$g(t) = \exp(tA)g(0) = e^{-\lambda t} \begin{bmatrix} 1 \\ \lambda t \\ \frac{(\lambda t)^2}{2} \\ \vdots \end{bmatrix}. \qquad (\exp(tA) = e^{-\lambda t} \exp(\lambda t I + tA))$$

Related distributions: Write $W_r = \min\{t : N_t \geq r\}$.

- Poisson distribution: $N_t \sim \text{Poisson}(\lambda t)$
- Continuous Uniform distribution: $W_1 | N_T = 1 \sim \text{Unif}(0, T)$
- Beta distribution: $\frac{1}{T} W_s | N_T = r \sim \text{Beta}(s, r - s + 1)$
- Multinomial distribution: $(N_{pT}, N_{qT} - N_{pT}, N_T - N_{qT}) | N_T = r \sim \text{Multi}(r, (p, q - p, 1 - q))$ (Trinomial)
- Exponential distribution: $W_1 \sim \text{Exp}(\frac{1}{\lambda})$
- Gamma distribution: $W_r \sim \text{Gamma}(r, \frac{1}{\lambda})$
- Beta distribution: $\frac{W_s}{W_r} \sim \text{Beta}(s, r - s)$ (can be generalized to Dirichlet distribution)

2 Exercises

2.1 김우철 (2016)

서울대학교 메일 계정에 수신되는 스팸메일의 수가 발생률 $\lambda_S = 2$(시간당)인 포아송 과정이다. 그리고 네이버 메일 계정에 수신되는 스팸메일의 수가 발생률 $\lambda_N = 1$(시간당)인 포아송 과정이며, 두 과정은 독립이다. 또한, V_k 는 서울대학교 메일 계정에 k 번째 스팸메일이 도착하기까지의 걸린 시간이고, W_k 는 네이버 메일 계정에 k 번째 스팸 메일이 도착하기까지의 걸린 시간이다.
(a) X 는 서울대학교 메일 계정에 오전 9시부터 저녁 6시까지 수신되는 스팸 메일의 수라고 정의하자. X 의 기댓값과 분산을 구하여라.
(b) $\mathbb{E}[V_{10}|V_2]$ 를 계산하여라.
(c) V_4/W_2 는 F 분포임을 밝히고, 그 모수값을 구하여라.
(d) $\mathbb{P}(V_2 > W_1)$ 을 계산하여라.

2.2 김우철 (2015, 2017)

발생률이 λ 인 포아송 과정 $\{N_t : t \geq 0\}$ 에서 r 번째 현상이 발생할 때까지의 시간을

$$W_r = \min\{t : N_t \geq r\} \qquad (r = 1, 2, \cdots)$$

이라고 할 때, 다음 물음에 답하여라.
(2015: a) W_r 과 N_t 의 관계를 이용하여 다음 등식이 성립하는 이유를 설명하여라.

$$\int_0^t \frac{\lambda^r}{\Gamma(r)} y^{r-1} e^{-\lambda y}\, dy = \sum_{k=r}^{\infty} \frac{e^{-\lambda t}(\lambda t)^k}{k!}$$

(2015: b) $\mathrm{Var}[\mathbb{E}(W_3 + W_4 + W_5|W_2)]$ 를 구하여라.
(2015: c) (W_1, W_2, W_3) 의 분산행렬을 구하여라.
(2015: d) W_1, W_2 의 일차함수 $aW_1 + bW_2 + c$ 로서

$$\mathbb{E}[(W_3 - (aW_1 + bW_2 + c))^2]$$

을 최소로 하는 a, b, c 를 구하여라.
(2015: e) $X = W_2/W_4, Y = W_4/W_5$ 라고 할 때 X 와 Y 의 결합확률밀도함수를 구하여라.
(2015: f) $T = (1 - 3X^2 + 2X^3)Y^4$ 의 확률밀도함수를 구하여라.
(2017: a) $X = W_1/W_2, Y = W_3/W_4$ 라고 할 때 X 와 Y 의 결합확률밀도함수를 구하여라.
(2017: b) $Z = XY^3$ 의 확률밀도함수를 구하여라.
(2017: c) $T = (4X - 1)^2$ 의 확률밀도함수를 구하여라.
(2017: d) $\mathrm{Cov}(N_{3t}, N_{5t}|N_t)$ 를 구하여라.
(2017: e) $\mathbb{E}[\mathrm{Cov}(N_t, N_{3t}|N_{5t})]$ 를 구하여라.

2.3 김우철 (2015, 2018)

서로 독립이고 성공률이 $0 < p < 1$ 인 베르누이 시행 $X_1, \cdots$ 을 관측하여 $r(= 1, 2, \cdots)$ 번째 성공까지의 시행횟수를 W_r 이라고 할 때 다음에 답하여라.
(2015: a) $\text{Cov}(W_1, W_3)$ 의 값을 구하여라.
(2015: b) $\text{Cov}(W_3, W_4|W_1)$ 의 값을 구하여라.
(2018: a) $W_2 = x$ 인 조건에서 $(W_3, W_4)^\top$ 의 조건부확률밀도함수 $pdf_{3,4|2}(y, z|x)$ 를 구하여라.
(2018: b) $\text{Cov}[\mathbb{E}(W_4|W_2), \mathbb{E}(W_6|W_2)]$ 와 $\mathbb{E}[\text{Cov}(W_4, W_6|W_2)]$ 를 구하여라.

2.4 김우철 (2018)

확률변수 $X_1, \cdots, X_k$ 가 서로 독립이고 각각 $\text{Poisson}(\lambda_i)$ 분포 ($i = 1, \cdots, k$)를 따르고,

$$N = X_1 + \cdots + X_k, X = (X_1, \cdots, X_k)^\top$$

라고 할 때 다음에 답하여라.
(a) $N = n$ 인 조건에서 X 의 조건부확률밀도함수 $pdf_{X|N}(x_1, \cdots, x_k|n)$ 을 구하여라.
(b) $\text{Var}[\mathbb{E}(X|N)]$ 과 $\mathbb{E}[\text{Var}(X|N)]$ 을 구하여라.

Wk 7 Exercises Answers

2.1 (a) $X \sim Poi(18)$ $\mathbb{E}(X) = 18$. $Var(X) = 18$.

(b) $V_{10} = V_2 + \tilde{V}_8$. $V_2 \perp\!\!\!\perp \tilde{V}_8$. $V_2 \sim Gamma(2, \frac{1}{2})$. $\tilde{V}_8 \sim Gamma(8, \frac{1}{2})$

$\mathbb{E}(V_{10} | V_2) = V_2 + \mathbb{E}(\tilde{V}_8) = V_2 + 4$.

(c) $V_4 \perp\!\!\!\perp W_2$. $V_4 \sim Gamma(4, \frac{1}{2})$, $W_2 \sim Gamma(2, 1)$.

$\Rightarrow 4V_4 \sim Gamma(4, 2) \overset{d}{=} \chi^2(8)$.

$2W_2 \sim Gamma(2, 2) \overset{d}{=} \chi^2(4)$.

$\therefore F = \frac{V_4}{W_2} = \frac{4V_4/8}{2W_2/4} \sim F(8, 4)$

(d) $V_2 \sim Gamma(2, \frac{1}{2}) \perp\!\!\!\perp W_1 \sim Gamma(1, 1)$.

$$\mathbb{P}(V_2 > W_1) = \int_0^\infty \int_0^{v_2} \frac{1}{\Gamma(2)(\frac{1}{2})^2} \frac{1}{\Gamma(1)\, 1^1} v_2 e^{-2v_2} e^{-w_1} \, dw_1 \, dv_2$$

$$= \int_0^\infty 4v_2 e^{-2v_2} (1 - e^{-v_2}) \, dv_2$$

$$= \int_0^\infty 4v_2 (e^{-2v_2} - e^{-3v_2}) \, dv_2 = \frac{4}{2^2} - \frac{4}{3^2} = \frac{5}{9}.$$

(2015)

2.2 (a) $\mathbb{P}(W_r \leq t) = \mathbb{P}(N_t \geq r)$.

(b) $\mathbb{E}(W_3 | W_2) = W_2 + \frac{1}{\lambda}$, $\mathbb{E}(W_4 | W_2) = W_2 + \frac{2}{\lambda}$, $\mathbb{E}(W_5 | W_2) = W_2 + \frac{3}{\lambda}$.

$Var(\mathbb{E}(W_3 + W_4 + W_5 | W_2)) = Var(3W_2 + \frac{6}{\lambda}) = 9\, Var(W_2) = \frac{18}{\lambda^2}$.

(c) Write $V_2 = W_2 - W_1$, $V_3 = W_3 - W_2$ so that W_1, V_2, V_3 mutually indep.

Since $\begin{bmatrix} W_1 \\ W_2 \\ W_3 \end{bmatrix} = \begin{bmatrix} 1 & & \\ 1 & 1 & \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} W_1 \\ V_2 \\ V_3 \end{bmatrix}$, one has $Var \begin{bmatrix} W_1 \\ W_2 \\ W_3 \end{bmatrix} = \begin{bmatrix} 1 & & \\ 1 & 1 & \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1/\lambda^2 & & \\ & 1/\lambda^2 & \\ & & 1/\lambda^2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ & 1 & 1 \\ & & 1 \end{bmatrix} = \frac{1}{\lambda^2} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}$

(d) **The best predictor (not necessarily linear) is the conditional expectation.**

That is, $\mathbb{E}(W_3 | W_1, W_2) = W_2 + \frac{1}{\lambda}$. Hence $(a, b, c) = (0, 1, \frac{1}{\lambda})$.

(e) Write $V_4 = W_4 - W_2$. $V_5 = W_5 - W_4$.

$$f_{W_2, V_4, V_5}(w_2, v_4, v_5) = \frac{\lambda^2}{\Gamma(2)} \frac{\lambda^2}{\Gamma(2)} \frac{\lambda^1}{\Gamma(1)} w_2 v_4 e^{-\lambda(w_2 + v_4 + v_5)} I_{(0,\infty)}(w_2) I_{(0,\infty)}(v_4) I_{(0,\infty)}(v_5).$$

Define $u: (w_2, v_4, v_5) \mapsto (x, y, z) = \left(\frac{w_2}{w_2 + v_4}, \frac{w_2 + v_4}{w_2 + v_4 + v_5}, w_2 + v_4 + v_5\right)$.

$u^{-1}: (x, y, z) \mapsto (xyz, (1-x)yz, (1-y)z)$

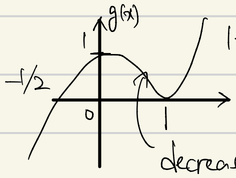
Compute the Jacobian det: $\left| \det \frac{\partial(w_2, v_4, v_5)}{\partial(x, y, z)} \right| = \left| \det \begin{bmatrix} yz & xz & xy \\ -yz & (1-x)z & (1-x)y \\ 0 & -z & 1-y \end{bmatrix} \right| = \left| z \cdot (y^2 z) + (1-y)(yz^2) \right| = yz^2$

Change of Variables: $f_{X,Y,Z}(x, y, z) = \lambda^5 (xyz)(1-x)yz e^{-\lambda z} yz^2 I_{(0,1)}(x) I_{(0,1)}(y) I_{(0,\infty)}(z)$.

Marginalize wrt z: $f_{X,Y}(x, y) = \left(6x(1-x) I_{(0,1)}(x)\right) \left(4y^3 I_{(0,1)}(y)\right)$ $\because \int_0^\infty \frac{\lambda^5}{24} z^4 e^{-\lambda z} dz = 1$.

(f) We've shown $f_X(x) = 6x(1-x) I_{(0,1)}(x) \perp\!\!\!\perp f_Y(y) = 4y^3 I_{(0,1)}(y)$

$T = (1-3X^2+2X^3)Y^4 = (1-X)^2(1+2X)Y^4$



$1-3x^2+2x^3 = g(x)$.

decreasing for $x \in (0,1)$. one-to-one.

Write $\tilde{X} = g(X) = (1-X)^2(1+2X)$.

$g'(x) = -6x(1-x)$. Define $g^{-1}: (0,1) \to (0,1)$.

Change of Variables: $f_{\tilde{X}}(\tilde{x}) = \dfrac{f_X(g^{-1}(\tilde{x}))}{|g'(g^{-1}(\tilde{x}))|} = 1 \cdot I_{(0,1)}(\tilde{x})$.

Write $\tilde{Y} = Y^4$.

Change of Variables: $f_{\tilde{Y}}(\tilde{y}) = \dfrac{f_Y(\sqrt[4]{\tilde{y}})}{|4(\sqrt[4]{\tilde{y}})^3|} = 1 \cdot I_{(0,1)}(\tilde{y})$.

Now, $T = \tilde{X}\tilde{Y}$. $\tilde{X}, \tilde{Y} \overset{iid}{\sim} \text{Unif}(0,1)$.

Finally, define $h: (\tilde{x}, \tilde{y}) \mapsto (t, \tilde{y}) = (\tilde{x}\tilde{y}, \tilde{y})$.

$h^{-1}: (t, \tilde{y}) \mapsto (t/\tilde{y}, \tilde{y})$

$$\left|\det \frac{\partial(\tilde{x}, \tilde{y})}{\partial(t, \tilde{y})}\right| = \left|\det \begin{bmatrix} 1/\tilde{y} & -t/\tilde{y}^2 \\ 0 & 1 \end{bmatrix}\right| = \frac{1}{\tilde{y}}.$$

$$\Rightarrow f_{T,\tilde{Y}}(t, \tilde{y}) = \frac{1}{\tilde{y}} I(0 < t < \tilde{y} < 1)$$ ← very important.

Marginalize with respect to $\tilde{y}$: $f_T(t) = \displaystyle\int_t^1 \frac{d\tilde{y}}{\tilde{y}} = (-\log t) I_{(0,1)}(t)$.

Another Shorter Solution.

$F_X(x) = 3x^2 - 2x^3$ for $x \in (0,1)$. $\Rightarrow$ $3X^2 - 2X^3 \sim \text{Unif}(0,1)$

$F_Y(y) = y^4$ for $y \in (0,1)$ $\Rightarrow$ $Y^4 \sim \text{Unif}(0,1)$.

(2017)

2.2 (a) $X = W_1/W_2$. $Y = W_3/W_4$.

Write $V_n = W_n - W_{n-1}$ for $n = 2,3,4$. $\Rightarrow X = \frac{W_1}{W_1+V_2}$, $Y = \frac{W_1+V_2+V_3}{W_1+V_2+V_3+V_4}$.

$f_{W_1,V_2,V_3,V_4}(w_1, v_2, v_3, v_4) = \lambda^4 e^{-\lambda(w_1+v_2+v_3+v_4)} I_{(0,\infty)^4}(w_1,v_2,v_3,v_4)$.

Define $u: (w_1, v_2, v_3, v_4) \longmapsto (x, \bar{x}, y, \bar{y}) = \left(\frac{w_1}{w_1+v_2}, \frac{w_1+v_2}{w_1+v_2+v_3}, \frac{w_1+v_2+v_3}{w_1+v_2+v_3+v_4}, w_1+v_2+v_3+v_4\right)$

$u^{-1}: (x, \bar{x}, y, \bar{y}) \longmapsto (x\bar{x}y\bar{y}, (1-x)\bar{x}y\bar{y}, (1-\bar{x})y\bar{y}, (1-y)\bar{y})$

$$\det \frac{\partial(w_1, v_2, v_3, v_4)}{\partial(x, \bar{x}, y, \bar{y})} = \det \begin{bmatrix} \bar{x}y\bar{y} & xy\bar{y} & x\bar{x}\bar{y} & x\bar{x}y \\ -\bar{x}y\bar{y} & (1-x)y\bar{y} & (1-x)\bar{x}\bar{y} & (1-x)\bar{x}y \\ 0 & -y\bar{y} & (1-\bar{x})\bar{y} & (1-\bar{x})y \\ 0 & 0 & -\bar{y} & 1-y \end{bmatrix}$$

$$= (\bar{x}y\bar{y})(y\bar{y})\bar{y} \det \begin{pmatrix} 1 & x & x\bar{x} & x\bar{x}y \\ -1 & 1-x & (1-x)\bar{x} & (1-x)\bar{x}y \\ 0 & -1 & 1-\bar{x} & (1-\bar{x})y \\ 0 & 0 & -1 & 1-y \end{pmatrix}$$

$$= \bar{x}y^2\bar{y}^3 \det \begin{bmatrix} 1 & x & x\bar{x} & x\bar{x}y \\ 0 & 1 & \bar{x} & \bar{x}y \\ 0 & 0 & 1 & y \\ 0 & 0 & 0 & 1 \end{bmatrix} = \bar{x}y^2\bar{y}^3.$$

Change of Variables: $f_{X,\bar{X},Y,\bar{Y}}(x,\bar{x},y,\bar{y}) = \lambda^4 e^{-\lambda\bar{y}} \bar{x}y^2\bar{y}^3 I_{(0,1)^3}(x,\bar{x},y) I_{(0,\infty)}(\bar{y})$.

All mutually indep.
$$\begin{cases} f_X(x) = I_{(0,1)}(x) & \sim Unif(0,1) \\ f_{\bar{X}}(\bar{x}) = 2\bar{x} I_{(0,1)}(\bar{x}) & \sim Beta(2,1) \\ f_Y(y) = 3y^2 I_{(0,1)}(y). & \sim Beta(3,1) \\ f_{\bar{Y}}(\bar{y}) = \frac{\lambda^4}{\Gamma(4)} e^{-\lambda\bar{y}} I_{(0,\infty)}(\bar{y}) & \sim Gamma(4, \tfrac{1}{\lambda}). \end{cases}$$

(b) $X, Y^3 \overset{iid}{\sim} Unif(0,1)$. $\Rightarrow Z = XY^3 \sim f_Z(z) = (-\log z) I_{(0,1)}(z)$. (See ⊕).

(c) $T = (4X-1)^2$. $X \sim Unif(0,1)$.

$$f_T(t) = \begin{cases} \frac{1}{4\sqrt{t}}, & t \in (0,1) \\ \frac{1}{8\sqrt{t}}, & t \in (1,9) \\ 0, & o.w. \end{cases}$$

(d) Write $\begin{cases} N_{3t} = N_t + M_{2t} \\ N_{5t} = N_{3t} + M'_{2t} \end{cases}$ $\Rightarrow$ N_t, M_{2t}, M'_{2t} mutually indep.

$$\Rightarrow \mathrm{Cov}(N_{3t}, N_{5t} \mid N_t) = \mathrm{Cov}(M_{2t}, M_{2t} + M'_{2t} \mid N_t)$$
$$= \mathrm{Cov}(M_{2t}, M_{2t} + M'_{2t})$$
$$= \mathrm{Var}(M_{2t})$$
$$= 2\lambda t.$$

(e) $(N_t, M_{2t}, M'_{2t}) \mid N_{5t} \sim \mathrm{Multi}(N_{5t}, (\frac{1}{5}, \frac{2}{5}, \frac{2}{5}))$.

$$\Rightarrow \mathrm{Var}(N_t, M_{2t}, M'_{2t} \mid N_{5t}) = N_{5t} \cdot \begin{bmatrix} \frac{1}{5}\cdot\frac{4}{5} & -\frac{1}{5}\cdot\frac{2}{5} & -\frac{1}{5}\cdot\frac{2}{5} \\ -\frac{1}{5}\cdot\frac{2}{5} & \frac{2}{5}\cdot\frac{3}{5} & -\frac{2}{5}\cdot\frac{2}{5} \\ -\frac{1}{5}\cdot\frac{2}{5} & -\frac{2}{5}\cdot\frac{2}{5} & \frac{2}{5}\cdot\frac{3}{5} \end{bmatrix} = \frac{1}{25} N_{5t} \begin{bmatrix} 4 & -2 & -2 \\ -2 & 6 & -4 \\ -2 & -4 & 6 \end{bmatrix}$$

Since $\begin{bmatrix} N_t \\ N_{3t} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} N_t \\ M_{2t} \\ M'_{2t} \end{bmatrix}$,

$$\mathrm{Var}(N_t, N_{3t} \mid N_{5t}) = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix} \cdot \frac{1}{25} N_{5t} \begin{bmatrix} 4 & -2 & -2 \\ -2 & 6 & -4 \\ -2 & -4 & 6 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} = \frac{1}{25} N_{5t} \begin{bmatrix} 4 & 2 \\ 2 & 6 \end{bmatrix}$$

$$\Rightarrow \mathrm{Cov}(N_t, N_{3t} \mid N_{5t}) = \frac{2}{25} N_{5t}.$$

$$\Rightarrow \mathbb{E}[\mathrm{Cov}(N_t, N_{3t} \mid N_{5t})] = \frac{2}{25} \mathbb{E}(N_{5t}) = \frac{2}{5}\lambda t.$$

2.3 (a) $\mathrm{Cov}(W_1, W_3) = \mathrm{Var}(W_1) = \frac{1-p}{p^2}$ $\because W_1 \sim \mathrm{Geo}(p)$.

(b) $\mathrm{Cov}(W_3, W_4 \mid W_1) = \mathrm{Cov}(W_2, W_3)$

$$= \mathrm{Var}(W_2) = 2 \cdot \frac{1-p}{p^2} \quad \because W_2 \sim NB(2, p).$$

(a) $f_{W_3, W_4 \mid W_2}(W_3 = y, W_4 = z \mid W_2 = x) = f_{W_3 \mid W_2}(y \mid x) f_{W_4 \mid W_3}(z \mid y)$

$$= (1-p)^{(y-x)-1} p \, I_{\{x+1, x+2, \cdots\}}(y) \cdot (1-p)^{(z-y)-1} p \, I_{\{y+1, y+2, \cdots\}}(z)$$

$$= (1-p)^{z-x-2} p^2 \, I(x < y < z < \infty, \ y, z \in \mathbb{N}).$$

(b) $\mathbb{E}(W_4 \mid W_2) = W_2 + \frac{2}{p}$, $\mathbb{E}(W_6 \mid W_2) = W_2 + \frac{4}{p}$, $\mathrm{Cov}(W_4, W_6 \mid W_2) = \mathrm{Cov}(W_2, W_4) = \mathrm{Var}(W_2) = 2 \cdot \frac{1-p}{p^2}$

$\Rightarrow \mathrm{Cov}[\mathbb{E}(W_4 \mid W_2), \mathbb{E}(W_6 \mid W_2)] = \mathrm{Var}(W_2) = 2 \cdot \frac{1-p}{p^2}$

$\mathbb{E}[\mathrm{Cov}(W_4, W_6 \mid W_2)] = 2 \cdot \frac{1-p}{p^2}$

irrelevant to W_2

2.4 (a) For $x_1, \cdots, x_k \in \mathbb{Z}_{\geq 0}$ s.t. $x_1 + \cdots + x_k = n$,

$$f_{X|N}(x_1, \cdots, x_k | n) = \frac{\prod_{i=1}^{k} \mathbb{P}(X_i = x_i)}{\mathbb{P}(N=n)} = \frac{\prod^{k} e^{-\lambda_i} \lambda_i^{x_i} / x_i!}{e^{-\lambda} \lambda^n / n!} = \binom{n}{x_1, x_2, \cdots, x_k} \left(\frac{\lambda_1}{\lambda}\right)^{x_1} \cdots \left(\frac{\lambda_k}{\lambda}\right)^{x_k} \text{ where } \lambda = \sum_{i=1}^{k} \lambda_i$$

Hence, $X|N \sim \text{Multi}(N, (\frac{\lambda_1}{\lambda}, \cdots, \frac{\lambda_k}{\lambda}))$

(b) $$\mathbb{E}(X|N) = \frac{N}{\lambda}\begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_k \end{bmatrix}, \quad \mathbb{V}ar(X|N) = \frac{N}{\lambda^2}\begin{bmatrix} \lambda_1(\lambda-\lambda_1) & -\lambda_1\lambda_2 & \cdots & -\lambda_1\lambda_k \\ -\lambda_1\lambda_2 & \lambda_2(\lambda-\lambda_2) & \cdots & -\lambda_2\lambda_k \\ \vdots & \vdots & \ddots & \vdots \\ -\lambda_1\lambda_k & -\lambda_2\lambda_k & \cdots & \lambda_k(\lambda-\lambda_k) \end{bmatrix}$$

and $\mathbb{E}(N) = \mathbb{V}ar(N) = \lambda$.

$$\therefore \text{Var}[\mathbb{E}(X|N)] = \frac{1}{\lambda}\begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_k \end{bmatrix} \lambda \cdot \frac{1}{\lambda}[\lambda_1 \cdots \lambda_k] = \frac{1}{\lambda}\begin{bmatrix} \lambda_1^2 & \lambda_1\lambda_2 & \\ & \ddots & \\ & & \lambda_k^2 \end{bmatrix}.$$

$$\text{and } \mathbb{E}[\mathbb{V}ar(X|N)] = \frac{1}{\lambda}\begin{bmatrix} \lambda_1(\lambda-\lambda_1) & -\lambda_1\lambda_2 & \cdots & -\lambda_1\lambda_k \\ -\lambda_1\lambda_2 & \lambda_2(\lambda-\lambda_2) & \cdots & -\lambda_2\lambda_k \\ \vdots & \vdots & \ddots & \vdots \\ -\lambda_1\lambda_k & -\lambda_2\lambda_k & \cdots & \lambda_k(\lambda-\lambda_k) \end{bmatrix}.$$

Rmk $\text{Var}[\mathbb{E}(X|N)] + \mathbb{E}[\mathbb{V}ar(X|N)] = \frac{1}{\lambda}\begin{bmatrix} \lambda_1\lambda & & 0 \\ & \ddots & \\ 0 & & \lambda_k\lambda \end{bmatrix} = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_k \end{bmatrix} = \mathbb{V}ar(X)$ as expected.

Midterm 2 Solution

1

Suppose $X_1, X_2 \sim \text{iid N}(0,1)$. Find the pdfs of

$$Y = \frac{\sigma X_1 + \mu X_2}{X_2}, \qquad Z = \frac{X_1 X_2}{\sqrt{X_1^2 + X_2^2}},$$

respectively.

1.1 Answer

(a) By the representative definition of the Cauchy destribution, one has $X_1/X_2 \sim \text{Cauchy}(0,1)$, i.e, the standard Cauchy distribution. Hence, scaling and translating give

$$f_Y(y) = \frac{1}{\pi\sigma\left(1 + \left(\frac{y-\mu}{\sigma}\right)^2\right)} \text{I}_{\mathbb{R}}(y).$$

One may write $Y \sim \text{Cauchy}(\mu, \sigma)$.
(b) Consider a map $u : (x_1, x_2) \mapsto (z, w)$ where

$$z = \frac{x_1 x_2}{\sqrt{x_1^2 + x_2^2}}, \qquad w = \frac{x_1^2 + x_2^2}{2}.$$

Note that u is a $4-1$ correspondence and that

$$\begin{aligned} \left|\det \frac{\partial(z,w)}{\partial(x_1,x_2)}\right| &= \left|\det \begin{pmatrix} \frac{x_2^3}{(x_1^2+x_2^2)^{3/2}} & \frac{x_1^3}{(x_1^2+x_2^2)^{3/2}} \\ x_1 & x_2 \end{pmatrix}\right| \\ &= \frac{|x_1^2 - x_2^2|}{\sqrt{x_1^2+x_2^2}} = \sqrt{2w - 4z^2}. \end{aligned}$$

By appealing to the Change of Variables method, one has

$$f_{Z,W}(z,w) = 4\frac{f_{X,Y}(x_1,x_2)}{\sqrt{2w-4z^2}} = \frac{\sqrt{2}}{\pi}\frac{e^{-w}}{\sqrt{w-2z^2}}\text{I}(w > 2z^2).$$

Integrate out w to attain the marginal pdf of Z.

$$\begin{aligned} f_Z(z) = \int_{2z^2}^{\infty} \frac{\sqrt{2}e^{-w}}{\pi\sqrt{w-2z^2}} dw &= \frac{\sqrt{2}e^{-2z^2}}{\pi}\int_0^\infty s^{-1/2}e^{-s} ds && (\text{let } s = w - 2z^2 > 0) \\ &= \sqrt{\frac{2}{\pi}} e^{-2z^2}. && (\Gamma(1/2) = \sqrt{\pi}) \end{aligned}$$

In fact, this is the pdf of $\text{N}(0, (1/2)^2)$.

2

Suppose $\{N_t : t \geq 0\}$ is a homogeneous Poisson process of rate λ. Define $W_r = \min\{t : N_t \geq r\}$ for each $r = 1, 2, \cdots$.
(a) Given positive integers $0 < k < l < m$, find $f_{Y_1|Y_2}(y_1|y_2)$ where $Y_1 = W_k/W_m, Y_2 = W_l/W_m$.
(b) Find $\mathbb{E}(Y_1Y_2)$. (c) Find $\mathrm{Cov}(Y_1, Y_2)$.

2.1 Answer

(a) Let $Z_2 = Y_2 - Y_1$. One has $(Y_1, Z_2) \sim \mathrm{Dir}(k, l-k, m-l)$. That is,

$$f_{Y_1,Z_2}(y_1, z_2) = \frac{\Gamma(m)}{\Gamma(k)\Gamma(l-k)\Gamma(m-l)} y_1^{k-1} z_2^{l-k-1}(1-y_1-z_2)^{m-l-1}\mathrm{I}_{\Delta^2}(y_1, z_2, 1-y_1-y_2),$$

where

$$\Delta^2 = \{(a,b,c) \in \mathbb{R}^3 : a, b, c > 0 = a+b+c-1\}.$$

Since 'shearing' $(y_1, z_2) \mapsto (y_1, y_1 + z_2) = (y_1, y_2)$ has the Jacobian determinant 1,

$$f_{Y_1,Y_2}(y_1, y_2) = \frac{\Gamma(m)}{\Gamma(k)\Gamma(l-k)\Gamma(m-l)} y_1^{k-1}(y_2-y_1)^{l-k-1}(1-y_2)^{m-l-1}\mathrm{I}(0 < y_1 < y_2 < 1).$$

It is obvious that Y_2 is marginally beta-distributed:

$$f_{Y_2}(y_2) = \frac{\Gamma(m)}{\Gamma(l)\Gamma(m-l)} y_2^{l-1}(1-y_2)^{m-l-1}\mathrm{I}(0 < y_2 < 1).$$

Dividing the two preceding equations yields the conditional pdf:

$$f_{Y_1|Y_2}(y_1|y_2) = \frac{\Gamma(l)}{\Gamma(k)\Gamma(l-k)} \frac{1}{y_2} \left(\frac{y_1}{y_2}\right)^{k-1} \left(1 - \frac{y_1}{y_2}\right)^{l-k-1} \mathrm{I}\left(0 < \frac{y_1}{y_2} < 1\right).$$

This shows that the conditional distribution $Y_1|Y_2$ is a scaled beta-distribution. That is,

$$\left.\frac{Y_1}{Y_2}\right| Y_2 \sim \mathrm{Beta}(k, l-k).$$

(b) By appealing to the law of iterated expectations,

$$\mathbb{E}(Y_1Y_2) = \mathbb{E}\left[\mathbb{E}\left(\left.\frac{Y_1}{Y_2}Y_2^2\right| Y_2\right)\right] = \mathbb{E}\left[\frac{k}{l}Y_2^2\right] = \frac{k}{l}\frac{l(l+1)}{m(m+1)} = \frac{k(l+1)}{m(m+1)}.$$

(c) Since $\mathbb{E}(Y_1) = \frac{k}{m}$ and $\mathbb{E}(Y_2) = \frac{l}{m}$, one has

$$\mathrm{Cov}(Y_1, Y_2) = \frac{k(l+1)}{m(m+1)} - \frac{k}{m}\frac{l}{m} = \frac{k}{m}\frac{m-l}{m(m+1)} = \frac{k(m-l)}{m^2(m+1)}.$$

3

Consider the following hierarchical models.
(a) Find Y where $Y|N \sim \text{Bin}(N,p)$ and $N \sim \text{Poi}(\lambda)$.
(b) Find $Y = \sum_{i=1}^n X_i$ where $X_i|p_i \sim \text{Ber}(p_i)$ and $p_i \sim$ iid $\text{Beta}(\alpha,\beta)$.
(c) Find Y where $Y|X \sim N(0,1/X)$ and $X \sim \text{Gamma}(\frac{n}{2},\frac{2}{n})$.

3.1 Answer

All we need is the law of total probability.
(a) Y is discrete. $Y \sim \text{Poi}(\lambda p)$ since

$$\begin{aligned} f_Y(y) = \sum_{n=0}^{\infty} f_{Y|N}(y|n)f_N(n) &= \sum_{n=y}^{\infty}\binom{n}{y}p^y(1-p)^{n-y}\frac{\lambda^n e^{-\lambda}}{n!} \\ &= \frac{(\lambda p)^y e^{-\lambda}}{y!}\sum_{n=y}^{\infty}\frac{(\lambda(1-p))^{n-y}}{(n-y)!} \\ &= \frac{(\lambda p)^y e^{-\lambda}}{y!}e^{\lambda(1-p)} = \frac{(\lambda p)^y e^{-\lambda p}}{y!}\mathrm{I}_{\{0,1,2,\cdots\}}(y). \end{aligned}$$

(b) Y is discrete. $Y \sim \text{Bin}(n,\alpha/(\alpha+\beta))$ since

$$\begin{aligned} f_Y(y) &= \sum_{x_1+\cdots+x_n=y}\int_{p_n\in[0,1]}\cdots\int_{p_1\in[0,1]}\prod_{i=1}^n f_{X_i|p_i}(x_i|p_i)f_{p_i}(p_i)dp_1\cdots dp_n \\ &= \sum_{x_1+\cdots+x_n=y}\int_{p_n\in[0,1]}\cdots\int_{p_1\in[0,1]}\prod_{i=1}^n p_i^{x_i}(1-p_i)^{1-x_i}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p_i^{\alpha-1}(1-p_i)^{\beta-1}dp_1\cdots dp_n \\ &= \sum_{x_1+\cdots+x_n=y}\prod_{i=1}^n\int_0^1\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p_i^{\alpha+x_i-1}(1-p_i)^{\beta+1-x_i-1}dp_i \\ &= \sum_{x_1+\cdots+x_n=y}\prod_{i=1}^n\frac{\Gamma(\alpha+\beta)\Gamma(\alpha+x_i)\Gamma(\beta+1-x_i)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha+\beta+1)} \\ &= \sum_{x_1+\cdots+x_n=y}\prod_{i=1}^n\left(\frac{\alpha}{\alpha+\beta}\right)^{x_i}\left(\frac{\beta}{\alpha+\beta}\right)^{1-x_i} \\ &= \binom{n}{y}\left(\frac{\alpha}{\alpha+\beta}\right)^{y}\left(\frac{\beta}{\alpha+\beta}\right)^{n-y}. \end{aligned}$$

(c) Y is continuous. $Y \sim t(n)$ since

$$\begin{aligned} f_y(y) = \int_0^{\infty} f_{Y|X}(y|x)f_X(x) &= \int_0^{\infty}\frac{x^{1/2}}{\sqrt{2\pi}}e^{-xy^2/2}\frac{(n/2)^{n/2}}{\Gamma(n/2)}x^{n/2-1}e^{-nx/2}dx \\ &= \frac{(n/2)^{n/2}}{\sqrt{2\pi}\Gamma(n/2)}\int_0^{\infty}x^{(n+1)/2-1}e^{-(y^2+n)x/2}dx \\ &= \frac{(n/2)^{n/2}}{\sqrt{2\pi}\Gamma(n/2)}\frac{\Gamma((n+1)/2)}{((y^2+n)/2)^{(n+1)/2}} = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)}\left(1+\frac{y^2}{n}\right)^{-(n+1)/2}. \end{aligned}$$

4

(a) Suppose $X \sim \text{Poi}(1)$. Find the values $a \in \{0, 1, 2, \cdots\}$ such that $\mathbb{E}[g_a(X)]$ exists where

$$g_a(x) = (x-a)!\mathrm{I}_{\{a,a+1,\cdots\}}(x).$$

(b) Suppose $X \sim \text{Gamma}(\alpha, 1)$ and $Y \sim \text{Poi}(x)$ for $\alpha \in \mathbb{N}$ and $x > 0$. Show that $\mathbb{P}(X \leq x) = \mathbb{P}(Y \geq \alpha)$.

4.1 Answer

(a) For $a = 2, 3, \cdots$, one has

$$\mathbb{E}[g_a(X)] = \sum_{x=a}^{\infty}(x-a)!\frac{e^{-1}}{x!} = \frac{e^{-1}}{a-1}\sum_{x=a}^{\infty}\left[\frac{(x-a)!}{(x-1)!} - \frac{(x+1-a)!}{(x+1-1)!}\right] = \frac{e^{-1}}{a-1}\frac{1}{(a-1)!} = \frac{e^{-1}}{a! - (a-1)!}.$$

It is obvious that $\mathbb{E}[g_a(X)] = \infty$ for $a = 0, 1$.
(b) Consider a homogeneous Poisson process of rate 1. Then X and Y represent W_α and N_x, respectively. $(W_\alpha \leq x)$ and $(N_x \geq \alpha)$ are the same events.

5

Suppose $X_1, \cdots, X_{n_1} \sim \text{iid N}(\mu_1, \sigma^2)$ and $Y_1, \cdots, Y_{n_2} \sim \text{iid N}(\mu_2, \sigma^2)$. Prove that

$$T = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{S_p\sqrt{n_1^{-1} + n_2^{-1}}} \sim t(n-2)$$

for the pooled variance $S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n-2}$ with $n = n_1 + n_2$.

5.1 Answer

By appealing to the normal sampling theory,

$$\overline{X} - \mu_1 \sim \text{N}(0, \frac{\sigma^2}{n_1}) \qquad (n_1-1)S_1^2/\sigma^2 \sim \chi^2(n_1-1)$$
$$\overline{Y} - \mu_2 \sim \text{N}(0, \frac{\sigma^2}{n_2}) \qquad (n_2-1)S_2^2/\sigma^2 \sim \chi^2(n_2-1)$$

and they are all independent. Hence,

$$Z = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sigma\sqrt{n_1^{-1} + n_2^{-1}}} \sim \text{N}(0, 1), \qquad V = (n-2)S_p^2/\sigma^2 \sim \chi^2(n-2)$$

are independent and $T = \frac{Z}{\sqrt{V/(n-2)}} \sim t(n-2)$ by the representative definition of the t-distribution.

6

Suppose $X_1, \cdots, X_n \sim \text{iid L}(0,1)$ for $n \geq 2$, i.e, $f(x) = \frac{e^x}{(1+e^x)^2}$. Find the pdf of

$$Y = \log\left(\frac{1+\exp(-X_{(1)})}{1+\exp(-X_{(n)})}\right)$$

where $X_{(r)}$ denotes the r-th order statistic for each $r = 1, 2, \cdots, n$.

6.1 Answer

The cdf is given by $F(x) = (1+e^{-x})^{-1}$. By appealing to the theory of probability integral transform,

$$F(X_{(r)}) \stackrel{d}{=} U_{(r)}$$

where $U_1, \cdots, U_n \sim \text{iid Unif}(0,1)$. Note that $Y = \log(F(X_{(n)})) - \log(F(X_{(1)})) \stackrel{d}{=} \log U_{(n)} - \log U_{(1)}$. Recall that $(U_{(1)}, U_{(n)})$ are jointly distributed by

$$f_{U_{(1)},U_{(n)}}(u_{(1)}, u_{(n)}) = \frac{n!}{(n-2)!}(u_{(n)} - u_{(1)})^{n-2}\mathrm{I}(0 < u_{(1)} < u_{(n)} < 1).$$

Now, consider a map

$$g : (u_{(1)}, u_{(n)}) \mapsto (v, y) = (-\log u_{(1)}, \log u_{(n)} - \log u_{(1)}).$$

The inverse g^{-1} is given by $(v, y) \mapsto (e^{-v}, e^{y-v})$ and hence the Jacobian determinant is given by

$$|\det dg^{-1}| = \left|\det\begin{pmatrix} -e^{-v} & 0 \\ -e^{y-v} & e^{y-v}\end{pmatrix}\right| = e^{y-2v}.$$

Apply the change of variables:

$$f_{V,Y}(v, y) = \frac{n!}{(n-2)!}(e^{y-v} - e^{-v})^{n-2}e^{y-2v}\mathrm{I}(0 < y < v < \infty).$$

Marginalize with respect to Y:

$$\begin{aligned} f_Y(y) &= \frac{n!}{(n-2)!}(e^y - 1)^{n-2}e^y \int_y^\infty e^{-nv}dv \\ &= (n-1)(e^y-1)^{n-2}e^y e^{-ny} \\ &= (n-1)e^{-y}(1-e^{-y})^{n-2}\mathrm{I}_{(0,\infty)}(y). \end{aligned}$$

As a remark, for $y > 0$,

$$f_Y(y) = \frac{d}{dy}(1-e^{-y})^{n-1}.$$

7

Suppose that a bus arrives at a bus stop following a homogeneous Poisson process of rate λ. Given time $T(>0)$, let W be the waiting time difference between the first and last passengers who arrived at the station. Compute $\mathbb{E}(W)$.

7.1 Answer

Assume total $N \geq 1$ passengers have arrived. If one writes the waiting times of the passengers by $W_1, \cdots, W_N$, then W_r is marginally beta-distributed for each $r = 1, \cdots, N$:

$$\frac{1}{T} W_r | N_T = N \sim \text{Beta}(r, N - r + 1).$$

That is, $\mathbb{E}(W_r | N_T = N) = \frac{r}{N+1} T$. By the linearity of expectation, one has

$$\mathbb{E}(W_N - W_1 | N_T = N) = \frac{N-1}{N+1} T.$$

Now observe that

$$W = \begin{cases} W_N - W_1, & N \geq 1 \\ 0, & N = 0 \end{cases}$$

and apply the law of iterated expectations:

$$\begin{aligned} \mathbb{E}(W) = \mathbb{E}[\mathbb{E}(W | N_T = N)] &= \mathbb{E}\left[\frac{N-1}{N+1} T \mathrm{I}(N \geq 1)\right] \\ &= \sum_{n=1}^{\infty} \frac{n-1}{n+1} T \frac{e^{-\lambda T} (\lambda T)^n}{n!} \\ &= \left(\sum_{n=1}^{\infty} \frac{(\lambda T)^n}{n!} - 2 \sum_{n=1}^{\infty} \frac{(\lambda T)^n}{(n+1)!} \right) T e^{-\lambda T} \\ &= \left(\sum_{n=1}^{\infty} \frac{(\lambda T)^n}{n!} - \frac{2}{\lambda T} \sum_{m=2}^{\infty} \frac{(\lambda T)^m}{m!} \right) T e^{-\lambda T} \\ &= \left((e^{\lambda T} - 1) - \frac{2}{\lambda T} \left(e^{\lambda T} - 1 - \lambda T\right) \right) T e^{-\lambda T} \\ &= (T - \frac{2}{\lambda}) + (T + \frac{2}{\lambda}) e^{-\lambda T}. \end{aligned}$$

- Please report any errors you find.

1 Multivariate(MVT) Normal Distribution

Bold characters $\mathbf{XYZ}$ denote vectors or matrices. $Usual\ characters\ XYZ$ denote scalars.

1.1 Characteristic Properties of MVT Normal Distribution

Definition 1 (MVT normal distribution: non-degenerate case)**.** *Let $\boldsymbol{\mu} \in \mathbb{R}^p$ and $\mathbf{\Sigma} > 0$. ($\mathbf{\Sigma}$ is a $p \times p$ real positive definite matrix.) A p-dimensional random vector $\mathbf{X}$ is defined to be (non-degenerate) normally-distributed if it admits a pdf*

$$f_{\mathbf{X}}(\mathbf{x}) = (\det(2\pi\mathbf{\Sigma}))^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}.$$

One writes $\mathbf{X} \sim \mathrm{N}_p(\boldsymbol{\mu}, \mathbf{\Sigma})$.

Proposition 1 (Translation and shaping)**.** $\mathbf{\Sigma}^{-1/2}(\mathbf{X}-\boldsymbol{\mu}) \sim \mathrm{N}_p(\mathbf{0}, \mathbf{I}_p)$ *where*

$$\mathbf{\Sigma} = \mathbf{\Sigma}^{1/2}(\mathbf{\Sigma}^{1/2})^\top$$

*is the **Cholesky Decomposition** of $\mathbf{\Sigma}$.*

Proposition 2 (Characteristic property I of the multivariate normal distribution)**.** *Suppose $\mathbf{X} \sim \mathrm{N}_p(\boldsymbol{\mu}, \mathbf{\Sigma})$. Define $Y = \mathbf{t}^\top\mathbf{X}$ provided that $\mathbf{t} \in \mathbb{R}^p$ is a p-dimensional vector. Then, $Y \sim \mathrm{N}_1(\mathbf{t}^\top\boldsymbol{\mu}, \mathbf{t}^\top\mathbf{\Sigma}\mathbf{t})$.*

- This can be proved via the moment generating function of $\mathbf{X}$. Note $\mathbb{E}(e^{\mathbf{t}^\top\mathbf{X}}) = \exp(\mathbf{t}^\top\boldsymbol{\mu} + \frac{1}{2}\mathbf{t}^\top\mathbf{\Sigma}\mathbf{t})$. Then, what is the mgf of $\mathbf{t}^\top\mathbf{X}$, namely, $\mathbb{E}(e^{s\mathbf{t}^\top\mathbf{X}})$ for real $s \in \mathbb{R}$? What does it imply?
- In fact, this characteristic property **defines** the multivariate normal distribution.

Definition 2 (MVT normal distribution: general case)**.** *Let $\boldsymbol{\mu} \in \mathbb{R}^p$ and $\mathbf{\Sigma} \geq 0$. ($\mathbf{\Sigma}$ is a $p \times p$ real positive **semi-**definite matrix.) A p-dimensional random vector $\mathbf{X}$ is defined to be normally-distributed if*

$$\mathbf{t}^\top\mathbf{X} \sim \mathrm{N}_1(\mathbf{t}^\top\boldsymbol{\mu}, \mathbf{t}^\top\mathbf{\Sigma}\mathbf{t})$$

*for **ALL** $\mathbf{t} \in \mathbb{R}^p$. If $\mathbf{\Sigma} > 0$ in addition, then the **Definitions 1 and 2** coincide. If $\mathbf{\Sigma} \not> 0$ on the contrary, then $\mathbf{\Sigma}$ is **NOT** invertible and $\mathbf{X}$ does **NOT** admit its pdf. Nevertheless, $\mathbf{X}$ is normally-distributed (degenerate case).*

Proposition 3 (Independence of multivariate normal distribution)**.** *Suppose $(\mathbf{X}, \mathbf{Y}) \sim$ multivariate normal distribution. $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}$ if and only if $\mathrm{Cov}(\mathbf{X}, \mathbf{Y}) = 0$.*

- (Question: replication) Suppose $X \sim \mathrm{N}(0, 1)$. Is $\mathbf{Y} = (X, X, X)$ normally-distributed?
- (Question: marginality) Let $\{\mathbf{e}_1, \cdots, \mathbf{e}_p\}$ be the standard (orthonormal) basis of $\mathbb{R}^p$. Suppose $\mathbf{X}$ is a p-dimensional random vector such that $\mathbf{e}_k^\top\mathbf{X}$ is normally-distributed for all $k = 1, \cdots, p$. Is $\mathbf{X}$ necessarily normally-distributed?
- (Question: independence) Suppose $X \sim \mathrm{N}(0, 1)$ and $Y \sim \mathrm{N}(0, 1)$ with $\mathrm{Cov}(X, Y) = 0$. Are X and Y necessarily independent?

Proposition 4 (Characteristic property II of the multivariate normal distribution)**.** *Suppose $\mathbf{X} \sim \mathrm{N}_p(\boldsymbol{\mu}, \mathbf{\Sigma})$. Define $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$ provided that $\mathbf{A}$ is a $q \times p$ real matrix and $\mathbf{b} \in \mathbb{R}^q$. Then, $\mathbf{Y} \sim \mathrm{N}_q(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\mathbf{\Sigma}\mathbf{A}^\top)$.*

1.2 True and False Implications

- Multivariate Normal $\Longleftrightarrow$ Jointly Normal $\Longrightarrow$ Marginally Normal $\not\Longrightarrow$ Jointly Normal
- Jointly Normal + Zero Covariance $\Longrightarrow$ Independence $\Longrightarrow$ Zero Covariance
- Marginally Normal + Zero Covariance $\not\Longrightarrow$ Independence
- Marginally Normal + Independence $\Longrightarrow$ Jointly Normal $\not\Longrightarrow$ Independence
- Key counterexample: $X \sim \mathrm{N}(0,1) \perp\!\!\!\perp Z \sim \mathrm{Ber}(1/2)$ and let $Y = (-1)^Z X$.
  Then, $Y \sim \mathrm{N}(0,1)$ and $\mathrm{Cov}(X,Y) = 0$ but $X \not\perp\!\!\!\perp Y$ and hence $(X,Y) \not\sim$ MVT Normal.

1.3 Quadratic Forms in Normal Random Vectors

Theorem 1 (Quadratic form through a real symmetric idempotent matrix). *Suppose* $\mathbf{X} \sim \mathrm{N}_p(\mathbf{0}, \sigma^2 \mathbf{I}_p)$ *and* $\mathbf{A}$ *is a* $p \times p$ *real symmetric idempotent matrix of rank* $m \le p$*. Then,*

$$Y = \frac{1}{\sigma^2} \mathbf{X}^\top \mathbf{A} \mathbf{X} \sim \chi^2(m).$$

- Note that $\mathbf{A}$ can be interpreted as a projection map onto an m-dimensional subspace V of $\mathbb{R}^p$ and that

$$(\textbf{degree of freedom}) = m = \mathrm{tr}\mathbf{A} = \mathrm{rank}\mathbf{A} = \dim \mathrm{im}\mathbf{A}.$$

- Moreover, $\mathbf{I} - \mathbf{A}$ is a projection map onto the orthogonoal complement $V^\perp$ of V.

$$p - m = \mathrm{tr}(\mathbf{I} - \mathbf{A}) = \mathrm{rank}(\mathbf{I} - \mathbf{A}) = \dim \mathrm{im}(\mathbf{I} - \mathbf{A}).$$

- (Example) For $p = n$, $\mathbf{A} = \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top$ represents a projection onto a (1-dimensional) line generated by $\mathbf{1}_n \in \mathbb{R}^n$. One the other hand, $\mathbf{I} - \mathbf{A} = \mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top$ represents a projection onto $\mathbf{1}_n^\perp$, which is $(n-1)$-dimensional subspace of $\mathbb{R}^n$. By some computation, one has

$$\mathbf{X} \overset{\mathbf{A}}{\mapsto} \overline{X} \mathbf{1}_n, \quad \frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{j=1}^n (X_j - \overline{X})^2 = \frac{1}{\sigma^2} \|(\mathbf{I} - \mathbf{A})\mathbf{X}\|^2 = \frac{1}{\sigma^2} \mathbf{X}^\top (\mathbf{I} - \mathbf{A}) \mathbf{X} \sim \chi^2(n-1).$$

Proposition 5. *Let* $\mathbf{X} \sim \mathrm{N}_p(\mathbf{0}, \mathbf{I}_p)$, $\mathbf{A}$ *be a* $p \times p$ *real symmetric matrix, and* $\mathbf{B}$ *be a* $k \times p$ *matrix. If* $\mathbf{BA} = 0$*, then* $\mathbf{BX}$ *and* $\mathbf{X}^\top \mathbf{A} \mathbf{X}$ *are independent.*

- Now, explain why $\overline{X} \perp\!\!\!\perp S^2$.

2 Exercises

2.1 김우철 (2016)

Suppose

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim \mathrm{N}_2\left(\begin{bmatrix} \mu \\ \mu \end{bmatrix}, \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix}\right).$$

Define $Y_1 = X_1 - X_2, Y_2 = X_1 + X_2$.
(a) (X_1, Y_1) 의 분포를 구하여라.
(b) Y_1 과 Y_2 는 독립임을 보여라.

2.2 이재용 (2020)

Suppose $\mathbf{Z} \sim \mathrm{N}_p(\mathbf{0}, \mathbf{I}_p)$. Let $\mathbf{A}$ be an arbitrary real symmetric idempotent $p \times p$ matrix. That is,

$$\mathbf{A}^2 = \mathbf{A} = \mathbf{A}^\top.$$

Prove that $\mathbf{Z}^\top \mathbf{A}\mathbf{Z} \sim \chi^2(\mathrm{tr}A)$. Now suppose $X_1, \cdots, X_n \sim \text{iid } \mathrm{N}(0, \sigma^2)$. Prove that

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{\sigma^2} \sim \chi^2(n-1).$$

2.3 김우철 (2015)

Consider the following linear regression model.

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \\ \mathbf{e} &\sim \mathrm{N}_n(0, \sigma^2\mathbf{V}) \end{aligned}$$

where $\mathbf{X}$ is a real $n \times (p+1)$ matrix of column full rank and $\mathbf{V}$ is a known $n \times n$ positive definite matrix. Assume $n > p+1$. Justify that

$$\begin{aligned} \widehat{\boldsymbol{\beta}} &= (\mathbf{X}^\top \mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top \mathbf{V}^{-1}\mathbf{y}, \\ \widehat{\sigma^2} &= (\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}})^\top \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}})/(n-p-1). \end{aligned}$$

Prove that

$$\frac{(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \mathbf{X}^\top \mathbf{V}^{-1}\mathbf{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{(p+1)\widehat{\sigma^2}} \sim \mathrm{F}(p+1, n-p-1).$$

2.3.1 Answer

- Justification of the estimators (Optional: Studied in Regression Analysis class)

Consider $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2)$. First we compute an MLE (Maximum Likelihood Estimator) $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}, \widehat{\sigma^2})$ for $\boldsymbol{\theta}$. The likelihood is given by

$$L(\boldsymbol{\theta}) = (\det(2\pi\sigma^2\mathbf{V}))^{-1/2} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right).$$

Take a negative logarithm:

$$-\log L(\boldsymbol{\theta}) = \frac{n}{2}\log(\sigma^2) + \frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + (\text{constant}). \tag{1}$$

We are to minimize (1) with respect to $\boldsymbol{\beta}, \sigma^2$. The equation (1) shows that minimizer $\widehat{\boldsymbol{\beta}}$ does not depend on the choice of σ^2. That is,

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta} \in \mathbb{R}^{p+1}}{\operatorname{argmin}} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \tag{2}$$

$$= (\mathbf{X}^\top \mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top \mathbf{V}^{-1}\mathbf{y}. \tag{3}$$

This can be justified in a number of ways. Firstly, one may take a derivative:

$$\frac{\partial}{\partial \boldsymbol{\beta}^\top}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = -2\mathbf{X}^\top \mathbf{V}^{-1}\mathbf{y} + 2\mathbf{X}^\top \mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta}.$$

Solving $\frac{\partial}{\partial \boldsymbol{\beta}^\top} = \mathbf{0}$ gives $\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top \mathbf{V}^{-1}\mathbf{y}$. Or if one writes $\widetilde{\mathbf{y}} = \mathbf{V}^{-1/2}\mathbf{y}$ and $\widetilde{\mathbf{X}} = \mathbf{V}^{-1/2}\mathbf{X}$, then

$$\begin{aligned}
\widehat{\boldsymbol{\beta}} &= \underset{\boldsymbol{\beta} \in \mathbb{R}^{p+1}}{\operatorname{argmin}} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\
&= \underset{\boldsymbol{\beta} \in \mathbb{R}^{p+1}}{\operatorname{argmin}} (\widetilde{\mathbf{y}} - \widetilde{\mathbf{X}}\boldsymbol{\beta})^\top (\widetilde{\mathbf{y}} - \widetilde{\mathbf{X}}\boldsymbol{\beta}) \\
&= (\widetilde{\mathbf{X}}^\top \widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^\top \widetilde{\mathbf{y}} \\
&= (\mathbf{X}^\top \mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top \mathbf{V}^{-1}\mathbf{y}.
\end{aligned}$$

Now we are to minimize (1) with respect to σ^2 given (3). Observe that

$$\underset{\sigma^2 > 0}{\operatorname{argmin}} \frac{n}{2}\log(\sigma^2) + \frac{C}{2\sigma^2} = \frac{C}{n}$$

and hence $\widehat{\sigma^2} = (\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}})^\top \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}})/n$. However, the MLE $\widehat{\sigma^2} = (\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}})^\top \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}})/n$ is biased. A simple correction suggests an unbiased estimator for σ^2:

$$\widehat{\sigma^2} = (\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}})^\top \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}})/(n - p - 1). \qquad \text{(So-called, MSE (Mean Squared Error))}$$

Unbiasedness of the estimator will be verified soon.

- Distributions of the estimator $\widehat{\boldsymbol{\theta}}$ (수리통계에서 배울 것이라고 회귀분석 교수님이 기대하는 바로 그것!)

We start from the fact that

$$\mathbf{y} \sim \mathrm{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{V}). \qquad \text{(That is, the regression model is well-specified.)}$$

Every linear map preserves normality of a random vector (**Proposition 4**). Therefore,

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{y} - \boldsymbol{\beta} \sim \mathrm{N}_{p+1}\left(\mathbb{E}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}), \mathrm{Var}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right)$$

where

$$\begin{aligned}
\mathbb{E}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= (\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{V}^{-1}(\mathbf{X}\boldsymbol{\beta}) - \boldsymbol{\beta} = \mathbf{0}, \\
\mathrm{Var}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= (\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{V}^{-1}(\sigma^2\mathbf{V})\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X})^{-1} = \sigma^2(\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X})^{-1}.
\end{aligned}$$

Then the **Proposition 1** ensures us that

$$\begin{aligned}
\frac{1}{\sigma}(\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X})^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &\sim \mathrm{N}_{p+1}(\mathbf{0}, \mathbf{I}_{p+1}), \\
\frac{1}{\sigma^2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &\sim \chi^2(p+1).
\end{aligned}$$

On the other hand, observe that

$$\mathbf{V}^{-1/2}(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) = \mathbf{A}\mathbf{V}^{-1/2}\mathbf{y} = \mathbf{A}\mathbf{V}^{-1/2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \tag{4}$$

(This step is not trivial at all. Please check it by yourself!)

where

$$\mathbf{A} = \mathbf{I} - \mathbf{V}^{-1/2}\mathbf{X}(\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{V}^{-1/2}$$

is an $n \times n$ real symmetric idempotent matrix. Define $\mathbf{Z} = \mathbf{V}^{-1/2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$. If the regression model is well-specified, then $\mathbf{Z} \sim \mathrm{N}_n(\mathbf{0}, \sigma^2\mathbf{I})$ and $\mathbf{V}^{-1/2}(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) = \mathbf{A}\mathbf{Z}$ by (4). It concludes that

$$\begin{aligned}
\frac{(n-p-1)\widehat{\sigma^2}}{\sigma^2} &= \frac{1}{\sigma^2}(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}})^\top\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) \\
&= \frac{1}{\sigma^2}(\mathbf{A}\mathbf{Z})^\top(\mathbf{A}\mathbf{Z}) \\
&= \frac{1}{\sigma^2}\mathbf{Z}^\top\mathbf{A}\mathbf{Z} \sim \chi^2\left(\mathrm{tr}(\mathbf{A})\right). \qquad \textbf{(Theorem 1)}
\end{aligned}$$

Commutativity of trace operator (i.e, $\mathrm{tr}(\mathbf{PQ}) = \mathrm{tr}(\mathbf{QP})$) proves that $\mathrm{tr}(\mathbf{A}) = n - p - 1$. We are almost done. It only remains to prove that $\widehat{\boldsymbol{\beta}} \perp\!\!\!\perp \widehat{\sigma^2}$. Recall that $\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{y}$ and that

$$(n-p-1)\widehat{\sigma^2} = \left\|\mathbf{A}\mathbf{V}^{-1/2}\mathbf{y}\right\|^2. \qquad \text{(See (4))}$$

By appealing to the **Proposition 3**, it suffices to show that $(\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{V}^{-1}(\sigma^2\mathbf{V})\mathbf{V}^{-1/2}\mathbf{A} = \mathbf{0}$. (why?)

1 L^r-spaces and Modes of Convergence

We are given a fixed probability space $(S, \mathcal{F}, \mathbb{P})$. Recall that a random variable X is a measurable function $X : S \to \mathbb{R}$ and that $\mathbb{E}(-) = \int_S (-) d\mathbb{P}$.

Definition 1. *Fix $0 < r < \infty$. L^r-space is said to be a space of random variables with finite r-th moments.*

$$L^r(S, \mathcal{F}, \mathbb{P}) = \{X : \mathbb{E}(|X|^r) < \infty\}$$

Definition 2. *Fix $r = \infty$. L^∞-space is said to be a space of random variables with finite essential supremums.*

$$\begin{aligned} L^\infty(S, \mathcal{F}, \mathbb{P}) &= \{X : \inf\{a \in \mathbb{R} : \mathbb{P}(X > a) = 0\} < \infty\} \\ &= \{X : \{a \in \mathbb{R} : \mathbb{P}(X > a) = 0\} \neq \emptyset\} \end{aligned}$$

Proposition 1. *By appealing to the* ***Lyapunov Inequality*** *previously described in the Chapter 1, one has*

$$L^1 \supsetneqq L^2 \supsetneqq \cdots \supsetneqq L^r \supsetneqq \cdots \supsetneqq L^\infty \supsetneqq \{X : \{\epsilon > 0 : \mathbb{E}(e^{tX}) < \infty, |t| < \epsilon\} \neq \emptyset\}$$

Fact 1. *Fix $1 \leq r \leq \infty$. L^r-space is a complete normed vector space, i.e,* ***Banach space****.*

$$\|X\|_{L^r} = \mathbb{E}(|X|^r)^{1/r}.$$

Fact 2. *Fix $r = 2$. L^2-space is a complete inner product vector space, i.e,* ***Hilbert space****.*

$$\langle X, Y \rangle_{L^2} = \mathbb{E}(XY), \|X\|_{L^2} = \sqrt{\langle X, X \rangle_{L^2}}.$$

Actually, we identify $X = X'$ if $\mathbb{P}(X = X') = 1$. Now fix r and consider a sequence $(X_n)_{n=1}^\infty$ of random variables in L^r-space. There are at least five modes of convergence in L^r-space. Of course, we only handle two of them in this course.

Definition 3 (Modes of Convergence)**.**

$$X_n(s) \to X(s) \forall s \in S \quad \text{(pointwise convergence)}$$
$$\mathbb{P}(\{s \in S : X_n(s) \to X(s)\}) = 1 \quad \text{(almost sure convergence)}$$
$$\mathbb{E}(|X_n - X|^r) \to 0 \quad \text{(convergence in norm)}$$
$$\mathbb{P}(\{s \in S : |X_n(s) - X(s)| > \epsilon\}) \to 0 \forall \epsilon > 0 \quad \text{(convergence in probability)}$$
$$\mathbb{P}(X_n < x) \to \mathbb{P}(X < x) \forall x \in \mathbb{R} \quad \text{(convergence in distribution)}$$

You only need to understand the last two concepts.

- Prove that $X_n \xrightarrow{p} X$ if and only if

$$\forall \epsilon > 0, \exists N, n > N \Longrightarrow \mathbb{P}(|X_n - X| > \epsilon) < \epsilon$$

The central topics in this course include the followings:

- WLLN
- CLT
- Δ-Method

2 Exercises: CLT and the Δ-method

2.1 이재용 (2016)

Suppose $X_1, \cdots, X_n \sim \text{iid Beta}(\alpha, 1)$ with $\alpha > 0$ and define $Y_n = \min X_i$. Find $r > 0$ such that $n^r Y_n$ admits a limiting distribution. Find the limiting distribution.

2.2 이재용 (2016)

Suppose $X_1, \cdots, X_n \sim \text{iid N}(0, \sigma^2)$ and $\sigma^2 > 0$. Prove that

$$\frac{\sum_{m=1}^n X_m}{\left(\sum_{m=1}^n X_m^2\right)^{1/2}} \xrightarrow{d} \text{N}(0,1).$$

Find the distribution of

$$Y = \frac{1}{1 + \sum_{m=1}^k X_m^2 / \sum_{m=k+1}^n X_m^2}.$$

2.3 이재용 (2016)

Suppose that

$$(X_1, Y_1), \cdots, (X_n, Y_n) \sim \text{N}\left(\begin{bmatrix}\mu_1 \\ \mu_2\end{bmatrix}, \begin{bmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{bmatrix}\right).$$

We have shown that $\sqrt{n}(\hat{\rho}_n - \rho) \xrightarrow{d} \text{N}(0, (1-\rho^2)^2)$ as $n \to \infty$ in the textbook. Now, find a function $g : (-1, 1) \to \mathbb{R}$ such that

$$\sqrt{n}(g(\hat{\rho}_n) - g(\rho)) \xrightarrow{d} \text{N}(0,1).$$

2.4 김우철 (2017)

Suppose $X_1, X_2, \cdots \sim \text{iid Ber}(p)$. Define the r-th waiting time by $W_r = \min\{n : \sum_{i=1}^n X_i \geq r\}$. Define

$$\hat{p}_r = \frac{r}{W_r}.$$

Find the limiting distribution of $\sqrt{r}(\hat{p}_r - p)$ as $r \to \infty$. Find a variance stabilizing transformation g such that

$$\sqrt{r}(g(\hat{p}_r) - g(p)) \xrightarrow{d} \text{N}(0,1).$$

2.5 김우철 (2015)

Suppose $U_{(1)} < U_{(2)} < \cdots < U_{(n)}$ are order statistics based on random samples from $\mathrm{Unif}(0,1)$. Define

$$R_n = \frac{U_{(1)}}{U_{(n)}}.$$

Find $r > 0$ such that $n^r R_n$ admits a limiting distribution. Find the limiting distribution.

1 Limiting Distribution = Asymptotic Distribution

아래의 기출문제들을 풀어보면 다음 두 가지 질문에 답할 수 있을 것이라 기대합니다:

- 수리통계 1의 5단원을 우리가 왜 배우는가?
  - **1.1** 문제 하나만 풀어봐도 5단원의 소중함을 느낄 수 있음. 이는 수리통계 2에 가서 더욱 두드러짐.
- 수리통계 1 기말고사는 왜 전 범위여야 하는가?(!)
  - 문제를 풀다보면 여러가지 확률변수 (3장), 변수변환, 균등분포 및 지수분포의 순서통계량 (4장) 개념의 사용은 불가피함. 예를 들어, "카이제곱분포는 중간고사 범위잖아요" makes no sense. 결국 5장은 1-4장을 기본으로 깔고 가는 수리통계 1의 climax이자, 수리통계 2의 시작점임.

1.1 이재용 (2016++)

Suppose $X_1, \cdots, X_n \sim \text{iid Beta}(\alpha, 1)$ with $\alpha > 0$ and define $Y_n = \min X_i$. ($n > 2$)
(a) Find $r > 0$ such that $n^r Y_n$ admits a limiting distribution. Find the limiting distribution.
(b) Prove that

$$\widehat{\alpha}_n = \frac{n-1}{-\sum_{i=1}^n \log X_i}$$

is an unbiased and consistent estimator of α.
(c) Show that

$$\sqrt{n}(\log \widehat{\alpha}_n - \log \alpha) \xrightarrow{d} \mathrm{N}(0,1)$$

as $n \to \infty$.

1.1.1 Answer

(a) Recall that $X \sim \text{Beta}(\alpha, 1)$ has a cdf given by

$$F_X(x) = \begin{cases} 1, & x \geq 1 \\ x^\alpha, & 0 \leq x < 1 \\ 0, & x < 0 \end{cases}$$

Now for $y \in (0,1)$, one has

$$\mathbb{P}(Y_n > y) = \prod_{i=1}^n \mathbb{P}(X_i > y) = (1 - y^\alpha)^n,$$

which implies that

$$\mathbb{P}(n^r Y_n \leq t) = 1 - \left(1 - \left(\frac{t}{n^r}\right)^\alpha\right)^n$$

holds for $t \in (0, n^r)$. If $r > 1/\alpha$, then for each fixed positive real $t > 0$,

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}(n^r Y_n \le t) &= \lim_{n\to\infty} 1 - \left(1 - \frac{t^\alpha}{n^{r\alpha}}\right)^{n^{r\alpha} n^{1-r\alpha}} \\ &= \lim_{n\to\infty} 1 - e^{-t^\alpha n^{1-r\alpha}} = 0,\end{aligned}$$

which concludes that $n^r Y_n$ does not admit its limiting distribution, i.e, diverges. (0 cannot be a cdf!) Note that the preceding argument makes sense solely because for every fixed positive real $t > 0$, it is guaranteed that $t \in (0, n^r)$ is true for sufficiently large n. On the other hand, if $r = 1/\alpha$, then for each fixed positive real $t > 0$,

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}(n^r Y_n \le t) &= \lim_{n\to\infty} 1 - \left(1 - \frac{t^\alpha}{n}\right)^{n} \\ &= 1 - e^{-t^\alpha},\end{aligned}$$

which concludes that $n^r Y_n \xrightarrow{d} A$ where A is defined to has a cdf given by

$$F_A(t) = \begin{cases} 1 - e^{-t^\alpha}, & t \ge 0 \\ 0, & t < 0 \end{cases}$$

As a remark, the distribution of A is called the **Weibull distribution** after Swedish mathematician Waloddi Weibull, who described it in detail in 1951. One may write $A \sim \mathrm{Weibull}(\alpha, 1)$. (No need to memorize) Finally, if $0 < r < 1/\alpha$, then it is a direct consequence of **Slutsky's Theorem** that

$$n^r Y_n = n^{r - \frac{1}{\alpha}} n^{\frac{1}{\alpha}} Y_n \xrightarrow{d} 0 \cdot A = 0.$$

(You may avoid applying the Slutsky's Theorem here. You may equivalently give a reason that $\lim_{n\to\infty} \mathbb{P}(n^r Y_n \le t) = 1$ for all $t > 0$.) In sum, one concludes that $n^r Y_n$ converges in distribution if and only if $0 < r \le 1/\alpha$ and that

$$n^r Y_n \xrightarrow{d} \begin{cases} \textit{diverges}, & r > 1/\alpha, \\ A \sim \mathrm{Weibull}(\alpha, 1), & r = 1/\alpha, \\ 0, & 0 < r < 1/\alpha \end{cases}$$

It suffices for tutees to define A by providing its cdf.
(b) Recall the notion of **probability integral transform**. One has $F_X(X) \stackrel{d}{=} U \sim \mathrm{Unif}(0, 1)$. This ensures us that $(X_i)^\alpha \stackrel{d}{=} U_i$ for each $i = 1, \cdots, n$ where $U_1, \cdots, U_n$ denote n iid random standard uniform samples. The notion also suggests that $-\log U_i \stackrel{d}{=} Z_i$ where $Z_1, \cdots, Z_n$ are n iid random standard exponential samples. As a consequence, one may write

$$\frac{\widehat{\alpha}_n}{\alpha} = \frac{n-1}{-\sum_{i=1}^n \log X_i^\alpha} \stackrel{d}{=} \frac{n-1}{-\sum_{i=1}^n \log U_i} \stackrel{d}{=} \frac{n-1}{\sum_{i=1}^n Z_i} \stackrel{d}{=} \frac{n-1}{V}$$

where $V \sim \text{Gamma}(n, 1)$. Now it remains to show that

$$\mathbb{E}\left(\frac{n-1}{V}\right) = 1 \quad \text{and} \quad \frac{n-1}{V} \xrightarrow{p} 1.$$

To begin with, V admits its pdf f_V defined by

$$f_V(v) = \frac{1}{\Gamma(n)} v^{n-1} e^{-v} \mathrm{I}_{(0,\infty)}(v).$$

In order to find the distribution of $W = 1/V$, consider a differentiable map $(0, \infty) \to (0, \infty) : v \mapsto w = 1/v$. **Change of variables** method gives us that

$$f_W(w) = \frac{1}{\Gamma(n)} w^{-(n-1)} e^{-1/w} \left| \frac{dv}{dw} \right| = \frac{1}{\Gamma(n)} w^{-(n+1)} e^{-1/w} \mathrm{I}_{(0,\infty)}(w).$$

As a remark, the distribution of W is called the **Inverse Gamma distribution**, which is very intuitive. One may write $W \sim \text{invGamma}(n, 1)$. (No need to memorize. Will be described in detail in the Bayesian course.) Hence, it is natural that

$$\int_0^\infty \frac{1}{\Gamma(n)} w^{-(n+1)} e^{-1/w} dw = \int_0^\infty \frac{1}{\Gamma(n-1)} w^{-n} e^{-1/w} dw = \int_0^\infty \frac{1}{\Gamma(n-2)} w^{-(n-1)} e^{-1/w} dw = 1.$$

The three integrands represent the pdfs of $\text{invGamma}(n, 1), \text{invGamma}(n-1, 1), \text{invGamma}(n-2, 1)$, resp. It is a direct result of the equalities that

$$\begin{aligned}
\mathbb{E}(W) &= \frac{\Gamma(n-1)}{\Gamma(n)} = \frac{1}{n-1}, \\
\mathbb{E}(W^2) &= \frac{\Gamma(n-2)}{\Gamma(n)} = \frac{1}{(n-1)(n-2)}, \\
\text{Var}(W) &= \frac{1}{(n-1)^2(n-2)}.
\end{aligned}$$

We have shown that $\mathbb{E}((n-1)W) = 1$ and that $\text{Var}((n-1)W) \to 0$ as $n \to \infty$. These end the proof.
(c) We have observed above that $\widehat{\alpha}_n/\alpha \stackrel{d}{=} (n-1)/V$ where $V = \sum_{i=1}^n Z_i$ and $Z_i \sim \text{iid Exp}(1)$. By appealing to the **Central Limit Theorem**, one has

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n Z_i - \mathbb{E}(Z_1)\right) \xrightarrow{d} \mathrm{N}(0, \text{Var}(Z_1)).$$

This may be rewritten as

$$\sqrt{n}\left(\frac{(n-1)\alpha}{n\widehat{\alpha}_n} - 1\right) \xrightarrow{d} \mathrm{N}(0, 1).$$

Apply the **Δ-method** for $g = -\log$; one has $(g'(1))^2 = 1$:

$$\sqrt{n}\left(\log\frac{n}{n-1} + \log\frac{\widehat{\alpha}_n}{\alpha} - 0\right) \xrightarrow{d} \mathrm{N}(0, 1).$$

The proof is done by the fact that $\sqrt{n}\log\frac{n}{n-1} \to 0$ as $n \to \infty$. The fact is obtained by the **Mean Value Theorem** (high school analysis):

$$\log\frac{n}{n-1} = \frac{\log n - \log(n-1)}{n-(n-1)} = \frac{1}{c_n} < \frac{1}{n-1}$$

for some $c_n \in (n-1, n)$.

1.2 이재용 (2016)

Suppose $X_1, \cdots, X_n \sim \text{iid N}(0, \sigma^2)$ and $\sigma^2 > 0$. Prove that

$$\frac{\sum_{m=1}^n X_m}{(\sum_{m=1}^n X_m^2)^{1/2}} \xrightarrow{d} \text{N}(0,1).$$

Find the distribution of

$$Y = \frac{1}{1 + \sum_{m=1}^k X_m^2 / \sum_{m=k+1}^n X_m^2}.$$

1.2.1 Answer

Apply the **CLT**:

$$\sqrt{n}\left(\frac{1}{n\sigma}\sum_{m=1}^n X_m - 0\right) \xrightarrow{d} \text{N}(0,1).$$

Apply the **WLLN**:

$$\frac{1}{n\sigma^2}\sum_{m=1}^n X_m^2 \xrightarrow{p} 1.$$

On the both sides, take $(-)^{-1/2}$, which is continuous at 1:

$$\sqrt{n}\sigma\left(\sum_{m=1}^n X_m^2\right)^{-1/2} \xrightarrow{p} 1.$$

Hence by the **Slutsky's Theorem**, one has

$$\frac{\sum_{m=1}^n X_m}{(\sum_{m=1}^n X_m^2)^{1/2}} = \sqrt{n}\sigma\left(\sum_{m=1}^n X_m^2\right)^{-1/2} \cdot \sqrt{n}\frac{1}{n\sigma}\sum_{m=1}^n X_m \xrightarrow{d} \text{N}(0,1).$$

On the other hand, let

$$V_1 = \sum_{m=k+1}^{n} X_m^2 \sim \chi^2(n-k) \overset{d}{=} \text{Gamma}\left(\frac{n-k}{2}, 2\right),$$
$$V_2 = \sum_{m=1}^{k} X_m^2 \sim \chi^2(k) \overset{d}{=} \text{Gamma}\left(\frac{k}{2}, 2\right).$$

As a result, one has

$$Y = \frac{V_1}{V_1 + V_2} \sim \text{Beta}\left(\frac{n-k}{2}, \frac{k}{2}\right).$$

1.3 이재용 (2016)

Suppose that

$$(X_1, Y_1), \cdots, (X_n, Y_n) \sim \text{N}\left(\begin{bmatrix}\mu_1 \\ \mu_2\end{bmatrix}, \begin{bmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{bmatrix}\right).$$

We have shown that $\sqrt{n}(\hat{\rho}_n - \rho) \overset{d}{\to} \text{N}(0, (1-\rho^2)^2)$ as $n \to \infty$ in the textbook. Now, find a function $g : (-1, 1) \to \mathbb{R}$ such that

$$\sqrt{n}(g(\hat{\rho}_n) - g(\rho)) \overset{d}{\to} \text{N}(0, 1).$$

1.3.1 Answer

By appealing to the **Δ-method**, it suffices to find g such that

$$(g'(\rho))^2 = \frac{1}{(1-\rho^2)^2}.$$

One possible answer is as follows:

$$g(\rho) = \frac{1}{2} \int \frac{1}{1-\rho} + \frac{1}{1+\rho}\, d\rho = \frac{1}{2} \log \frac{1+\rho}{1-\rho}.$$

g is called a **variance stabilizing transformation**. As a remark, if g is a variance stabilizing transformation, then so is

$$h = \pm g + C,$$

where C is an arbitrary constant. Conventionally, however, we choose g that is increasing.

1.4 김우철 (2017)

Suppose $X_1, X_2, \cdots \sim \text{iid Ber}(p)$. Define the r-th waiting time by $W_r = \min\{n : \sum_{i=1}^n X_i \geq r\}$. Define

$$\widehat{p}_r = \frac{r}{W_r}.$$

Find the limiting distribution of $\sqrt{r}(\hat{p}_r - p)$ as $r \to \infty$. Find a variance stabilizing transformation g such that

$$\sqrt{r}(g(\hat{p}_r) - g(p)) \xrightarrow{d} \text{N}(0, 1).$$

1.4.1 Answer

Define $V_i = W_i - W_{i-1}$ for each $i = 1, \cdots, r$ with $W_0 = 0$. Then $V_1, \cdots, V_r \sim \text{iid Geo}(p)$ satisfy $W_r = V_1 + \cdots + V_r$. By the **CLT**, one has

$$\sqrt{r}\left(\frac{W_r}{r} - \mathbb{E}(V_1)\right) \xrightarrow{d} \text{N}(0, \text{Var}(V_1))$$

and hence

$$\sqrt{r}\left(\frac{1}{\widehat{p}_r} - \frac{1}{p}\right) \xrightarrow{d} \text{N}\left(0, \frac{1-p}{p^2}\right).$$

Apply the **Δ-method** to obtain

$$\sqrt{r}(\widehat{p}_r - p) \xrightarrow{d} \text{N}\left(0, p^2(1-p)\right).$$

Now it suffices to find g such that $(g'(p))^2 = 1/(p^2(1-p))$. One possible answer is as follows:

$$g(p) = \int \frac{dp}{p\sqrt{1-p}} = \int \frac{-2udu}{(1-u^2)u} = -\int \frac{1}{1-u} + \frac{1}{1+u}\,du = \log\frac{1-u}{1+u} = \log\frac{1-\sqrt{1-p}}{1+\sqrt{1-p}}.$$

(Substitute u for $\sqrt{1-p}$.)

1.5 김우철 (2015++)

Suppose $U_{(1)} < U_{(2)} < \cdots < U_{(n)}$ are order statistics based on random samples from $\text{Unif}(0, 1)$. Define

$$R_n = \frac{U_{(1)}}{U_{(n)}}.$$

(a) Find $s > 0$ such that $n^s R_n$ admits a limiting distribution. Find the limiting distribution.
(b) Prove that $U_{(n)} \xrightarrow{p} 1$ as $n \to \infty$. Find the limiting distribution of $n(1 - U_{(n)})$.
(c) Find the pdf of

$$Y = \frac{(U_{(r+1)})^r}{U_{(1)} \cdots U_{(r)}}$$

for each $1 < r < n$.

1.5.1 Answer

(a) Recall that for $r = 1, \cdots, n$, one has

$$-\log U_{(n-r+1)} \overset{d}{=} Z_{(r)} \overset{d}{=} \frac{V_1}{n} + \frac{V_2}{n-1} + \cdots + \frac{V_r}{n-r+1}$$

where $Z_1, \cdots, Z_n \sim \text{iid Exp}(1), V_1, \cdots, V_n \sim \text{iid Exp}(1)$, and $Z_{(1)} < \cdots < Z_{(n)}$ are order statistics. Now

$$-\log R_n = -\log U_{(1)} + \log U_{(n)} \overset{d}{=} \frac{V_2}{n-1} + \cdots + \frac{V_n}{1}$$

shows us that $R_n \perp\!\!\!\perp U_{(n)}$ and that $R_n \overset{d}{=} \tilde{U}_{(1)}$ where $\tilde{U}_{(1)} < \cdots < \tilde{U}_{(n-1)}$ are order statistics based on $(n-1)$ random uniform samples. That is, $R_n \sim \text{Beta}(1, n-1)$. Hence, for $t \in (0, n^s)$, one has

$$\mathbb{P}(n^s R_n \leq t) = \int_0^{t/n^s} (n-1)(1-x)^{n-2}\, dx = 1 - (1 - \frac{t}{n^s})^{n-1}.$$

By a similar argument to **Exercise 1.1**, one concludes that

$$n^s R_n \xrightarrow{d} \begin{cases} diverges, & s > 1 \\ \text{Exp}(1), & s = 1 \\ 0, & 0 < s < 1 \end{cases}$$

(b) Recall that $U_{(n)} \sim \text{Beta}(n, 1)$. Fix an arbitrarily small positive real $\epsilon > 0$. One has

$$\mathbb{P}\left(|U_{(n)} - 1| > \epsilon\right) = \mathbb{P}\left(U_{(n)} < 1 - \epsilon\right) = (1-\epsilon)^n \to 0$$

as $n \to \infty$. This concludes that $U_{(n)} \xrightarrow{p} 1$ by definition. Furthermore, verify that

$$\mathbb{P}\left(n(1 - U_{(n)}) \leq t\right) = \mathbb{P}\left(U_{(n)} \geq 1 - \frac{t}{n}\right) = 1 - \left(1 - \frac{t}{n}\right)^n \to 1 - e^{-t}$$

holds for $t > 0$. That is, $n(1 - U_{(n)}) \xrightarrow{d} \text{Exp}(1)$.
(c) Recall that

$$\begin{aligned} -\log U_{(1)} &\overset{d}{=} \frac{V_1}{n} + \frac{V_2}{n-1} + \cdots + \frac{V_n}{1}, \\ &\vdots \\ -\log U_{(r)} &\overset{d}{=} \frac{V_1}{n} + \frac{V_2}{n-1} + \cdots + \frac{V_{n-r+1}}{r}, \\ -\log U_{(r+1)} &\overset{d}{=} \frac{V_1}{n} + \frac{V_2}{n-1} + \cdots + \frac{V_{n-r}}{r+1}. \end{aligned}$$

Therefore, for $i = 1, \cdots, r$, it follows out that

$$-\log U_{(i)} + \log U_{(r+1)} \overset{d}{=} \frac{V_{n-r+1}}{r} + \cdots + \frac{V_{n-i+1}}{i},$$

and that

$$-\log U_{(i)} + \log U_{(r+1)} \overset{d}{=} \tilde{Z}_{(r-i+1)},$$

where $\tilde{Z}_{(1)} < \cdots < \tilde{Z}_{(r)}$ are order statistics based on r random standard exponential samples. Finally,

$$\log Y = \sum_{i=1}^{r} \left(-\log U_{(i)} + \log U_{(r+1)}\right) \overset{d}{=} \sum_{i=1}^{r} \tilde{Z}_{(r-i+1)} = \sum_{i=1}^{r} \tilde{Z}_i \overset{d}{=} X \sim \mathrm{Gamma}(r, 1).$$

It only remains to compute the pdf of $Y = e^X$ where $X \sim \mathrm{Gamma}(r, 1)$. Consider the exponential map $\exp : (0, \infty) \to (1, \infty)$ and apply the **Change of variables** to it.

$$f_Y(y) = f_X(\log y) \left| \frac{dx}{dy} \right| = \frac{1}{\Gamma(r)} (\log y)^{r-1} e^{-\log y} \frac{1}{y} = \frac{1}{\Gamma(r)} y^{-2} (\log y)^{r-1} \mathrm{I}_{(1,\infty)}(y).$$

1.6 Unknown

Suppose $X_1, \cdots, X_n \sim$ iid $\mathrm{Poi}(\mu)$ with $\mu > 0$. Find a variance stabilizing transformation g such that

$$\sqrt{n}\left(g(\overline{X}_n) - g(\mu)\right) \xrightarrow{d} \mathrm{N}(0, 1).$$

1.6.1 Answer

Thanks to the **CLT**, one has

$$\sqrt{n}\left(\overline{X}_n - \mu\right) \xrightarrow{d} \mathrm{N}(0, \mu).$$

It suffices to find g such that $(g'(\mu))^2 = 1/\mu$. One possible answer is $g(\mu) = 2\sqrt{\mu}$.

1.7 Unknown

Suppose that $\mathbf{Y} \sim \mathrm{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$ where $\boldsymbol{\beta} : p \times 1, \sigma^2 > 0$. Assume that $\mathbf{X}$ is a known $n \times p$ matrix and that $\mathbf{X}^\top\mathbf{X}$ is non-singular.
(a) Find the distribution of

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}.$$

(b) Let $\Pi = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$ and $SSE = \mathbf{Y}^\top(\mathbf{I} - \Pi)\mathbf{Y}$. Show that

$$SSE/\sigma^2 \sim \chi^2(n-p).$$

(c) Are $\widehat{\boldsymbol{\beta}}$ and SSE are independent? Answer with reasoning.
(d) Let $\widehat{\sigma^2} = SSE/(n-p)$. Find the distribution of

$$F = \frac{1}{p}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \mathbf{X}^\top \mathbf{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})/\widehat{\sigma^2}.$$

1.7.1 ANSWER

Duplicate to **Exercise 2.3, Week 9**.