VE401 REVIEW CLASS
MID 1

June 3. 2018

Announcement

- Midterm1:
  1. Time: 2018/6/6, Wed, 12:10-13:50
  2. Location: Dong Shang Yuan 415 and Dong Shang Yuan 412
  3. No cheating sheet
  4. Calculators not able to be connected to the Internet allowed

- No regular RC and OH this week
- Mid1 Office Hour:
  1. Time: 6.5(Tue) 20:00-22:00
  2. Location: JI New Building 326B

Materials covered: Part I

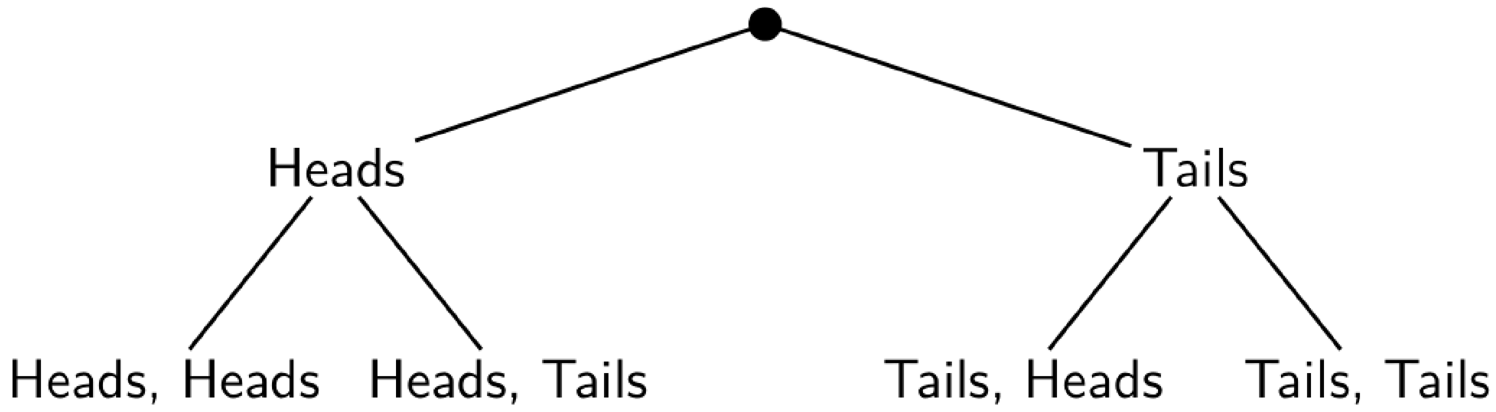
- Introduction to Probability and Counting
- Discrete Random Variables
- Continuous Random Variables
- Joint Distributions

Introduction to Probability and Counting

Probability calculation problem:

- Counting
- Conditional probability
- Total Probability

Counting by tree diagram



Basic Principles of Counting

- Permutation of k objects: $\frac{n!}{(n-k)!}$
- Combination of k objects from n objects: $\frac{n!}{k!(n-k)!}$
- Partition n objects into k disjoint subsets: $\frac{n!}{n_1!n_2!\ldots n_k!}$

Probability calculation problem:

- Counting
- Conditional probability
- Total Probability

Conditional Probability

1.1.16. Definition. Let $A, B \subset S$ be events and $P[A] \neq 0$. Then we define the conditional probability

$$P[B \mid A] := \frac{P[A \cap B]}{P[A]}.$$

- Example on Slide 46

Bayes's Theorem

1.1.21. Bayes's Theorem. Let $A_1, \dots, A_n \subset S$ be a set of mutually exclusive events whose union is S. Let $B \subset S$ be any event such that $P[B] \neq 0$. Then for any A_k, $k = 1, \dots, n$,

$$P[A_k \mid B] = \frac{P[B \cap A_k]}{P[B]} = \frac{P[B \mid A_k] \cdot P[A_k]}{\sum_{j=1}^{n} P[B \mid A_j] \cdot P[A_j]}.$$

- Example on Slide 59(disease problem)
- Example on Slide 61(the Monty Hall Paradox)

Probability calculation problem:

- Counting
- Conditional probability
- Total Probability

Total Probability

1.1.20. Theorem. Let $A_1, \ldots, A_n \subset S$ be a set of mutually exclusive events whose union is S. Let $B \subset S$ be any event. Then

$$P[B] = \sum_{j=1}^{n} P[B \mid A_j] \cdot P[A_j]. \tag{1.1.3}$$

The expression (1.1.3) is called the ***total probability*** formula for $P[B]$.

- Example on slide 53(the marriage problem)

Proof questions

- Make full use of all the conditions and assumptions in the problem.
- Think about how you can use these conditions if you have no idea about how to start the proof.

Properties of Probability

$$P[S] = 1, \qquad P[\emptyset] = 0, \qquad P[A^c] = 1 - P[A],$$

$$P[A_1 \cup A_2] = P[A_1] + P[A_2] - P[A_1 \cap A_2].$$

$$P[A_1 \cup A_2] \leq P[A_1] + P[A_2]$$

$$\text{if } A_1 \subset A_2, \text{ then } P[A_1] \leq P[A_2].$$

Independence of Events

1.1.18. Definition. Let A, B be two events. We say that A and B are independent if

$$P[A \cap B] = P[A]P[B]. \tag{1.1.2}$$

Equation (1.1.2) is equivalent to

$$\begin{aligned} P[A \mid B] &= P[A] \qquad &&\text{if } P[B] \neq 0, \\ P[B \mid A] &= P[B] \qquad &&\text{if } P[A] \neq 0. \end{aligned}$$

- Example on slide 49(the same birthday problem)

Exercise 1:

- Some investigation shows 5% of males and 0.25% of females have difficulty telling colors. Now we pick a person from a group of people that have same numbers of males and females. The person turns out to have some color blindness. So what’s the probability that the person is male?

Exercise 1:

- Some investigation shows 5% of males and 0.25% of females have difficulty telling colors. Now we pick a person from a group of people that have same numbers of males and females. The person turns out to have some color blindness. So what’s the probability that the person is male?

A: 95.24%

Exercise 2:

- If P(A) > 0, prove: $P(B \mid A) \geq 1 - \dfrac{P(\overline{B})}{P(A)}$

Exercise 2:

- If P(A) > 0, prove: $P(B \mid A) \geq 1 - \frac{P(\overline{B})}{P(A)}$

A:

$$
\begin{aligned}
& P(\mathrm{B} \mid \mathrm{A}) \\
& = \frac{P(\mathrm{AB})}{P(\mathrm{A})} \\
& = \frac{P(\mathrm{A}) + \mathrm{P(B)} - \mathrm{P(A+B)}}{P(\mathrm{A})} \\
& \geq \frac{P(\mathrm{A}) + \mathrm{P(B)} - 1}{P(\mathrm{A})} \\
& = \frac{P(\mathrm{A}) - P(\overline{B})}{P(\mathrm{A})} = 1 - \frac{P(\overline{B})}{P(\mathrm{A})}
\end{aligned}
$$

Discrete Random Variables

Discrete Random Variables

1.2.2. Definition. Let S be a sample space and Ω a countable subset of $\mathbb{R}$. A ***discrete random variable*** is a map

$$X : S \to \Omega$$

together with a function

$$f_X : \Omega \to \mathbb{R}$$

having the properties that

(i) $f_X(x) \geq 0$ for all $x \in \Omega$ and

(ii) $\sum_{x \in \Omega} f_X(x) = 1$.

The function f_X is called the ***probability density function*** or ***probability distribution*** of X.

Bernoulli Random Variable

1.2.3. Definition. Let S be a sample space and

$$X\colon S \to \{0,1\} \subset \mathbb{R}.$$

Let $0 < p < 1$ and define the density function

$$f_X\colon \{0,1\} \to \mathbb{R}, \qquad f_X(x) = \begin{cases} 1-p & \text{for } x = 0 \\ p & \text{for } x = 1. \end{cases}$$

We say that X is a ***Bernoulli random variable*** or follows a ***Bernoulli distribution*** with parameter p. We indicate this by writing

$$X \sim \text{Bernoulli}(p)$$

Cumulative Distribution Function

In practice, we are also often interested in the ***cumulative distribution function*** of a random variable, defined as follows,

$$F_X : \mathbb{R} \to \mathbb{R}, \qquad F_X(x) := P[X \leq x]$$

For a discrete random variable

$$F_X(x) = \sum_{y \leq x} f_X(y)$$

Note:
The function FX is a CDF if and only if the following three conditions hold:

1. $\lim_{x\to-\infty} F_X(x) = 0$ and $\lim_{x\to\infty} F_X(x) = 1$.
2. F_X is non-decreasing.
3. $\forall x_0 \in \mathbb{R} \quad \lim_{x\to x_0^+} = F_X(x_0)$

Relation between CDF and PDF

$$F_X(x) = \sum_{\xi \le x} f_X(x) \quad \Longleftrightarrow \quad f_X(x) = F_X(x) - F_X(x-1)$$

Expectation

1.2.7. Definition. Let (X, f_X) be a discrete random variable. Then the expected value of X is

$$\mathrm{E}[X] = \sum_{x \in \Omega} x \cdot f_X(x).$$

provided that the sum (possibly series if Ω is infinite) on the right converges absolutely.

Variance

1.2.12. Definition. Let X be a random variable with expectation $\mathrm{E}[X]$. Then the ***variance*** of X is defined as the ***mean square deviation from the mean***,

$$\mathrm{Var}\, X := \mathrm{E}\big[(X - \mathrm{E}[X])^2\big].$$

This type of ***mean square deviation*** will play an important role in many contexts of statistics.

1.2.13. Notation. For short (and especially in statistics) we write

$$\mathrm{E}[X] = \mu_X = \mu, \qquad \mathrm{Var}\, X = \sigma_X^2 = \sigma^2.$$

Standard Deviation

1.2.14. Definition. Let X be a random variable with variance σ_X^2. Then we define the ***standard deviation*** of X by

$$\sigma_X = \sqrt{\operatorname{Var} X} = \sqrt{\sigma_X^2}.$$

Properties

1.2.9. Corollary. By taking $H(x) = c \in \mathbb{R}$ (constant) and $H(x) = c \cdot x$ for some $c \in \mathbb{R}$, we immediately obtain

$$\mathrm{E}[c] = c, \qquad \mathrm{E}[cX] = c\,\mathrm{E}[X]$$

1.2.10. Theorem. Let X and Y be random variables. Then

$$\mathrm{E}[X+Y] = \mathrm{E}[X] + \mathrm{E}[Y].$$

1.2.15. Lemma. Let X be a random variable and $c \in \mathbb{R}$ a constant. Then

$$\operatorname{Var} c = 0, \qquad \operatorname{Var} cX = c^2 \operatorname{Var}[X]$$

1.2.16. Theorem. Let X and Y be independent random variables. Then

$$\operatorname{Var}[X+Y] = \operatorname{Var}[X] + \operatorname{Var}[Y].$$

The Moment-Generating Function

1.2.21. Definition. Let (X, f_X) be a random variable and $\mathrm{E}[X^k]$ the k^{th} ordinary moment of X. If the power series

$$m_X(t) := \sum_{k=0}^{\infty} \frac{\mathrm{E}[X^k]}{k!} t^k$$

has radius of convergence $\varepsilon > 0$, the thereby defined function

$$m_X(t)\colon (-\varepsilon, \varepsilon) \to \mathbb{R}$$

is called the ***moment-generating function*** for X.

Properties

1.2.22. Theorem. The moment-generating function exists if and only $\mathrm{E}[e^{tX}]$ exists, in which case

$$m_X(t) = \mathrm{E}[e^{tX}].$$

Furthermore,

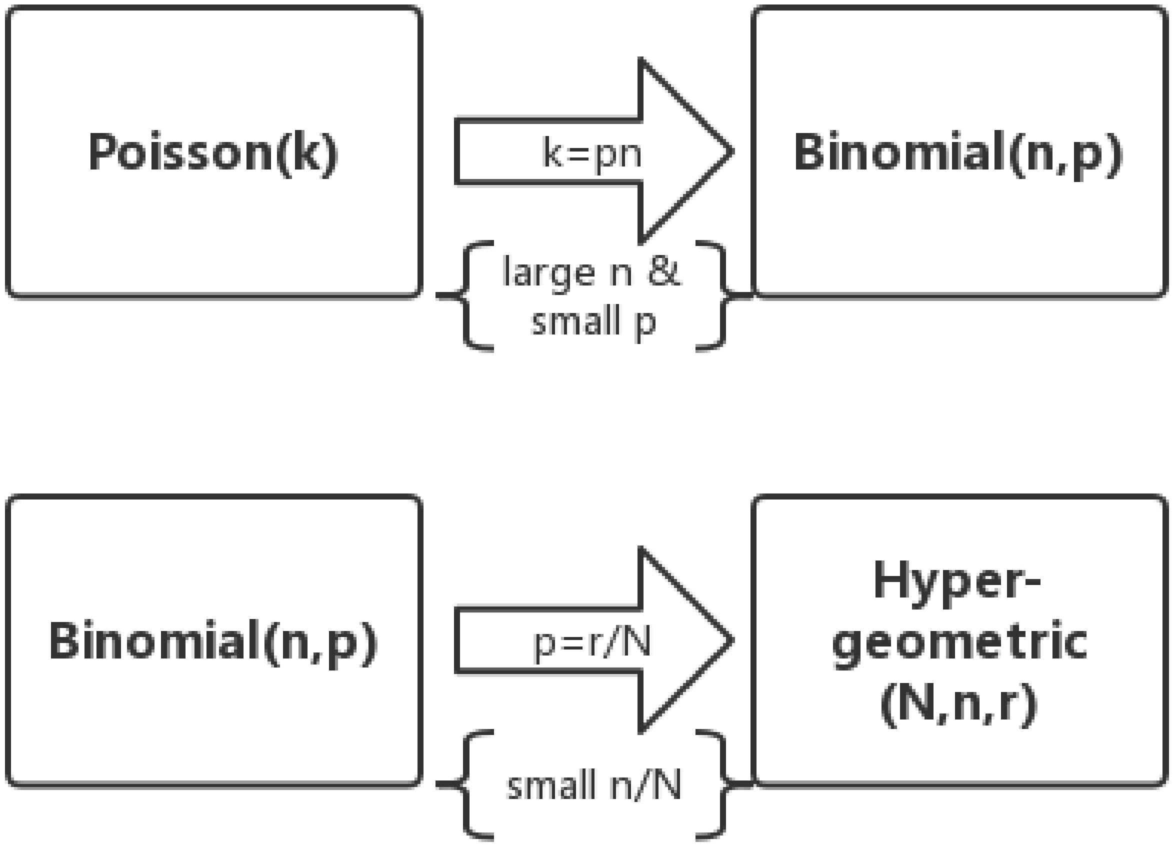
$$\mathrm{E}[X^k] = \left.\frac{d^k m_X(t)}{dt^k}\right|_{t=0}.$$

Summary of Discrete Distribution

Type(parameters)	Ω	$f_X(x)$	$m_X(t)$	$\mathbb{E}[X]$	Var X
Bernoulli(p)	$\{0, 1\}$	$p^x(1-p)^{1-x}$	$1-p+pe^t$	p	$p(1-p)$
Uniform(n)	$[1, n] \cap \mathbb{N}$	n^{-1}	$\frac{e^t(1-e^{nt})}{n(1-e^t)}$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$
Geometric(p)	$\mathbb{N}^*$	$(1-p)^{x-1}p$	$\frac{pe^t}{1-(1-p)e^t}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Binomial(n, p)	$[0, n] \cap \mathbb{N}$	$\binom{n}{x}p^x(1-p)^{n-x}$	$(1-p+pe^t)^n$	np	$np(1-p)$
Pascal(r, p)	$[r, +\infty) \cap \mathbb{N}$	$\binom{x-1}{r-1}p^r(1-p)^{x-r}$	$\frac{(pe^t)^r}{[1-(1-p)e^t]^r}$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$
Poisson(k)	$\mathbb{N}$	$\frac{k^x e^{-k}}{x!}$	$e^{k(e^t-1)}$	k	k
Hypergeometric(n, r, N)	$[\max(0, n-(N-r)), \min(n, r)] \cap \mathbb{N}$	$\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$		$\frac{nr}{N}$	$\frac{nr}{N}\frac{N-r}{N}\frac{N-n}{N-1}$

1. The parameter p in the table always satisfy 0<p<1
2. The moment generating function of hypergeometric distribution will not be discussed in this course.
3. Be aware of the domain of the moment generating function.

Two approximatings



Exercise 1

The log-distribution with parameter $p \in (0,1)$ is defined to have the following distribution

$$\mathbb{P}(X = x) = -\frac{(1-p)^x}{x \log p} \quad x \in \mathbb{N}^*$$

1. Prove that $\mathbb{P}(X \in \mathbb{N}^*) = 1$.
2. Calculate its expectation and variance.
3. Find its moment generating function.

Solution 1

1.

$$P(X \in N^*)$$
$$= \sum_{x=1}^{\infty} - \frac{(1-p)^x}{x \ln p}$$
$$= -\frac{1}{\ln p} \sum_{x=1}^{\infty} \frac{q^x}{x}$$
$$= -\frac{1}{\ln p} \int_0^q \frac{d}{dq}\left(\sum_{x=1}^{\infty} \frac{q^x}{x}\right) dq$$
$$= -\frac{1}{\ln p} \int_0^q \sum_{x=1}^{\infty} \frac{d}{dq}\left(\frac{q^x}{x}\right) dq$$
$$= -\frac{1}{\ln p} \int_0^q \sum_{x=1}^{\infty} q^{x-1} dq$$
$$= -\frac{1}{\ln p} \int_0^q \frac{1-q \cdot q^{\infty}}{1-q} dq$$
$$= -\frac{1}{\ln p} \cdot (-\ln(1-q))$$
$$= 1$$

2.

$$E[X]$$
$$= \sum_{x=1}^{\infty} x \cdot \frac{-q^x}{x \ln p}$$
$$= \frac{-1}{\ln p} \sum_{x=1}^{\infty} q^x$$
$$= \frac{-1}{\ln p} \frac{q}{p}$$

$$E[X^2]$$
$$= \sum_{x=1}^{\infty} - \frac{x q^x}{\ln p}$$
$$= -\frac{1}{\ln p} \sum_{x=1}^{\infty} x q^x$$
$$= \frac{-q}{p^2 \ln p}$$

$$Var[X]$$
$$= E[X^2] - E[X]^2$$
$$= -\frac{1}{\ln p} \frac{q}{p^2}\left(1 + \frac{q}{\ln p}\right)$$

3.

$$m_X(t)$$
$$= \sum_{x=1}^{\infty} - e^{tx} \cdot \frac{(1-p)^x}{x \ln p}$$
$$= -\frac{1}{\ln p} \sum_{x=1}^{\infty} \frac{(qe^t)^x}{x}$$
$$= -\frac{1}{\ln p} \int_0^z \frac{d}{dz}\left(\sum_{x=1}^{\infty} \frac{z^x}{x}\right) dz$$
$$= -\frac{1}{\ln p} \int_0^z \sum_{x=1}^{\infty} \frac{d}{dz}\left(\frac{z^x}{x}\right) dz$$
$$= -\frac{1}{\ln p} \int_0^z \sum_{x=1}^{\infty} z^{x-1} dz$$
$$= -\frac{1}{\ln p} \int_0^z \frac{1-z \cdot z^{\infty}}{1-z} dz$$
$$= -\frac{1}{\ln p} \cdot (-\ln(1-z))$$
$$= \frac{\ln(1-qe^t)}{\ln p}$$

where p+q=1,

z=qe^t

Exercise 2

Demonstrate that the sum of two independent Poisson Random Variables are still a Poisson Random Variable. Namely, if $X \sim Poisson(\lambda)$ and $Y \sim Poisson(\mu)$, where X and Y are independent. Show that $Z \sim Poisson(\lambda+\mu)$, where $Z=X+Y$.

Solution 2

Solution: Using the *total probability* formula, we have

$$
\begin{aligned}
\Pr(X+Y=k) &= \sum_{i=0}^{k} \Pr(X+Y=k \mid X=i)\cdot\Pr(X=i) && (12)\\
&= \sum_{i=0}^{k} \Pr(i+Y=k \mid X=i)\cdot\Pr(X=i) && (13)\\
&= \sum_{i=0}^{k} \Pr(Y=k-i)\Pr(X=i) && (14)\\
&= \sum_{i=0}^{k} e^{-\mu}\frac{\mu^{k-i}}{(k-i)!}e^{-\lambda}\frac{\lambda^i}{i!} && (15)\\
&= e^{-(\mu+\lambda)}\frac{1}{k!}\sum_{i=0}^{k}\frac{k!}{i!(k-i)!}\mu^{k-i}\lambda^i && (16)\\
&= e^{-(\mu+\lambda)}\frac{1}{k!}\sum_{i=0}^{k}\binom{k}{i}\mu^{k-i}\lambda^i && (17)\\
&= e^{-(\mu+\lambda)}\cdot\frac{(\mu+\lambda)^k}{k!} && (18)
\end{aligned}
$$

Thus $X+Y \sim \text{Poisson}\,(\lambda+\mu)$.

Or consider using the mgf. We know that $m_X(t)=\mathbb{E}[e^{tX}]=e^{\lambda(e^t-1)}$, and $m_Y(t)=\mathbb{E}[e^{tY}]=e^{\mu(e^t-1)}$, thus by independence of X and Y,

$$m_{X+Y}(t)=\mathbb{E}[e^{t(X+Y)}]=\mathbb{E}[e^{tX}e^{tY}]=\mathbb{E}[e^{tX}]\mathbb{E}[e^{tY}]=m_X(t)m_Y(t) \qquad (19)$$

and therefore

$$m_{X+Y}(t)=e^{\lambda(e^t-1)}e^{\mu(e^t-1)}=e^{(\lambda+\mu)(e^t-1)} \qquad (20)$$

Thus $X+Y \sim \text{Poisson}\,(\lambda+\mu)$.

Derivation of approximating Hypergeometric by Binomial

- http://www.m-hikari.com/ams/ams-2014/ams-13-16-2014/teerapabolarnAMS13-16-2014.pdf
- https://ac.els-cdn.com/S0378375899001871/1-s2.0-S0378375899001871-main.pdf?_tid=07db84a0-70f6-43a2-b0ca-949e85dc7d83&acdnat=1527950630_c312921b2657485658775c65c0b0a59e

Continuous Random Variable

1. Definition (p131)

2. $P[a \leq X \leq b] = \int_a^b f_X(x)dx$

   For any specific x, $P_X(x) = \int_x^x f_X(y)dy = 0$

3. Cumulative distribution:

   $F_X(x) \coloneqq P[X \leq x] = \int_{-\infty}^{x} f_X(y)dy$

4. Expectation:

   $E[X] = \int_{\mathbb{R}} x \cdot f_X(x)dx$

5. Variance: the same as that of discrete random variable

6. Definition of expectation, variance (p134)

Distribution

	P	f_X	m_X	E	Var
Exponential	$\beta \in \mathbb{R}$ $\beta > 0$	$\begin{cases} \frac{1}{\beta} e^{-x/\beta}, & x > 0, \\ 0, & x \leq 0, \end{cases}$	$(-\infty, \frac{1}{\beta}) \to \mathbb{R}$ $(1-\beta t)^{-1}$	β	β^2
Gamma	$\alpha, \beta \in \mathbb{R}$ $\alpha, \beta > 0$	$\begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, & x > 0, \\ 0, & x \leq 0, \end{cases}$	$(-\infty, \frac{1}{\beta}) \to \mathbb{R}$ $(1-\beta t)^{-\alpha}$	$\alpha\beta$	$\alpha\beta^2$
Normal	$\mu \in \mathbb{R}$ $\sigma > 0$	$\frac{1}{\sqrt{2\pi}\sigma} e^{-((x-\mu)/\sigma)^2/2}$	$\mathbb{R} \to \mathbb{R}$ $e^{\mu t + \sigma^2 t^2/2}$	μ	σ^2
Standard Normal	$\mu = 0$ $\sigma = 1$	$\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$	$\mathbb{R} \to \mathbb{R}$ $e^{t^2/2}$	0	1
Weibull (two-parameter)	$\alpha, \beta > 0$	$\begin{cases} \alpha\beta x^{\beta-1} e^{-\alpha x^\beta}, & x > 0, \\ 0, & x \leq 0, \end{cases}$		μ_W	σ_W^2
Uniform		$\begin{cases} 1, & 0 \leq x \leq 1, \\ 0, & otherwise \end{cases}$			

Note:

- $\Gamma(\alpha) = \int_0^\infty z^{\alpha-1} e^{-z}, \alpha > 0$: Euler gamma function.
- $\Gamma(1) = 1, \Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$ if $\alpha > 1$.
- $\mu_W = \alpha^{-1/\beta}\Gamma(1 + 1/\beta), \sigma_W^2 = \alpha^{-2/\beta}\Gamma(1 + 2/\beta) - \mu^2$.

Meanings of distributions:

- Exponential: the time between successive arrivals in a Poisson process
- Gamma: the time for j arrivals in a Poisson process
  - Chi-squared distribution (special case): when $\alpha = \frac{\gamma}{2}, \beta = 2$

$$f_{\chi_\gamma^2}(x) = \frac{1}{2^{\frac{\gamma}{2}}\Gamma\left(\frac{\gamma}{2}\right)} x^{\frac{\gamma}{2}-1} e^{-\frac{x}{2}}$$

- Normal: a continuous form of binomial distribution (personal understanding)

Transformation of random variables

- $$f_Y(y) = \begin{cases} f_X(\varphi^{-1}(y)) \cdot \left|\frac{d\varphi^{-1}(y)}{dy}\right|, & y \in \operatorname{ran} \varphi \\ 0, & y \notin \operatorname{ran} \varphi \end{cases}$$

- Transform from normal distribution to standard normal distribution

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \qquad Z = \frac{X - \mu}{\sigma}$$

Approximation: Normal $\mu,\ \sigma \longrightarrow$ Binomial n, p

- Conditions:
  - $p \approx 0.5, n > 10;$
  - $p < 0.5, n(1-p) > 5$
  - $p > 0.5, np > 5$
- Approximated by $\mu = np, \sigma^2 = np(1-p)$
  - Half unit correction

Approximating the Binomial Distribution

Hence, for $y = 0, \dots, n$,

$$P[X \le y] = \sum_{x=0}^{y} \binom{n}{x} p^x (1-p)^{n-x} \approx \Phi\left(\frac{y + 1/2 - np}{\sqrt{np(1-p)}}\right).$$

This additional term $1/2$ is known as the ***half-unit correction*** for the normal approximation to the cumulative binomial distribution function.

Chebyshev's Inequality (p174)

$P[-k\sigma < X - \mu < k\sigma] \geq 1 - \frac{1}{k^2}$ or $P[|X - \mu| \geq k\sigma] \leq \frac{1}{k^2}$

Reliability

- Failure density $f(t)$
- Reliability function

$$R(t) = 1 - \int_0^t f(x)dx = 1 - F(t)$$

- Hazard rate

$$\rho(x) = \frac{f(x)}{R(x)} = \frac{F'(x)}{R(x)} = -\frac{R'(x)}{R(x)}$$

- System
  - Series:

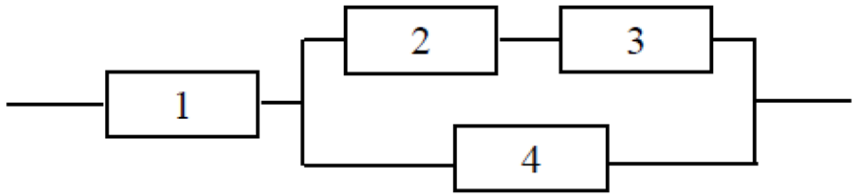
$$R_S(t) = \prod_{i=1}^{k} R_i(t)$$

  - Parallel:

$$R_P(t) = 1 - \prod_{i=1}^{k} (1 - R_i(t))$$

Exercise

1. Get the reliability of the following systems: the vales of the four independent components are p_1, p_2, p_3, p_4.



$$p_1[1-(1-p_2p_3)(1-p_4)] = p_1p_2p_3 + p_1p_4 - p_1p_2p_3p_4$$

Exercise

2. k follows uniform distribution in (0,5), then what's the probability that the solutions of $4x^2 + 4kx + k + 2 = 0$ are in $\mathbb{R}$?

$(4K)^2 - 4 \times 4 \times (K+2) \geq 0$，得到：$K \leq -1$ 或 $K \geq 2$

$$P(K \leq -1 or K \geq 2) = \frac{3}{5}$$

Exercise

3. The waiting time X of a customer in the bank follows exponential distribution with $\beta = 5$. If the total waiting time exceeds 10 minutes, he will leave. Suppose that the customer need to go to the bank five times a month, denote the times he leave without being served as Y, write the distribution function of Y, and calculate the value of $P(Y \geq 1)$.

$$P(X > 10) = 1 - P(X \leq 10) = 1 - \int_0^{10} \frac{1}{5} e^{-\frac{x}{5}} dx = e^{-2}$$

$$P(Y \geq 1) = 1 - P(Y = 0) = 1 - \binom{0}{5} (e^{-2})^0 (1 - e^{-2})^5 \approx 0.5167$$

Exercise

4. Points fall randomly on the edge of a circle with radius R. The center of the circle is the origin point of cartesian coordinates. What's the variance of the x-coordinate of the points?

$$F(x)=\begin{cases} 0, x<-R \\ \dfrac{\arccos\frac{-x}{R}}{\pi}, -R\le x<0 \\ 1-\dfrac{\arccos\frac{x}{R}}{\pi}, 0\le x<R \\ 1, x\ge R \end{cases}$$

$$f(x)=\begin{cases} \dfrac{1}{\pi R}\dfrac{1}{\sqrt{1-(\frac{x}{R})^2}}, -R\le x\le R \\ 0, other \end{cases}$$

$$E(x)=\frac{1}{\pi R}\int_{-R}^{R}\frac{Rx}{\sqrt{R^2-x^2}}dx=0$$

$$D(x)=\frac{1}{\pi R}\int_{-R}^{R}\frac{Rx^2}{\sqrt{R^2-x^2}}dx=\frac{R^2}{2}$$

Joint Distributions

Discrete Bivariate Random Variable

1.4.1. Definition. Let S be a sample space and Ω a subset of $\mathbb{Z}^2$. A discrete ***bivariate random variable*** is a map $(X, Y)\colon S \to \Omega$ together with a function $f_{XY}\colon \Omega \to \mathbb{R}$ with the properties that

(i) $f_{XY}(x, y) \geq 0$ for all $(x, y) \in \Omega$ and

(ii) $\sum_{(x,y)\in\Omega} f_{XY}(x, y) = 1$.

Then f_{XY} gives the probability that the pair (X, Y) assumes a given value (x, y), i.e.,

$$f_{XY}(x, y) = P[X = x \text{ and } Y = y].$$

The function f_{XY} is called the joint density function of the random variable (X, Y).

Marginal Density

1.4.4. Definition. Let $((X, Y), f_{XY})$ be a discrete bivariate random variable. We define the ***marginal density*** f_X for X by

$$f_X(x) = \sum_y f_{XY}(x, y)$$

and the marginal density f_Y for Y analogously.

Marginal Density

For discrete random variables X, Y, if we have known $f_X(x)$ and $f_Y(y)$, can we determine $f_{XY}(x, y)$ uniquely?

No! See Ex.1.

Continuous Bivariate Random Variables

1.4.6. Definition. Let S be a sample space. A continuous bivariate random variable is a map $(X, Y)\colon S \to \mathbb{R}^2$ together with a function $f_{XY}\colon \mathbb{R}^2 \to \mathbb{R}$ (called the ***joint density function***) with the properties that

(i) $f_{XY}(x, y) \geq 0$ for all $(x, y) \in \mathbb{R}^2$ and

(ii) $\int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} f_{XY}(x, y)\, dy\, dx = 1.$

The integral of f_{XY} is interpreted as the probability that X and Y assume values (x, y) in a given range, i.e.,

$$P[a \leq X \leq b \text{ and } c \leq Y \leq d] = \int_a^b \int_c^d f_{XY}(x, y)\, dy\, dx$$

for $a \leq b$, $c \leq d$. More generally, for any $\Omega \subset \mathbb{R}^2$,

$$P[(X, Y) \in \Omega] = \iint_{\Omega} f_{XY}(x, y)\, d(x, y).$$

Marginal Density

1.4.7. Definition. Let $((X, Y), f_{XY})$ be a continuous bivariate random variable. We define the ***marginal density*** f_X for X by

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y)\, dy$$

and the marginal density f_Y for Y analogously.

Independence

1.4.8. Definition. Let $((X, Y), f_{XY})$ be a bivariate random variable with marginal densities f_X and f_Y. If

$$\operatorname{dom} f_{XY} = (\operatorname{dom} f_X) \times (\operatorname{dom} f_Y)$$

and

$$f_{XY}(x, y) = f_X(x) f_Y(y) \qquad \text{for all } (x, y) \in \operatorname{dom} f_{XY}$$

then (X, f_X) and (Y, f_Y) are independent random variables.

Conditional Densities

1.4.10. Definition. Let $((X, Y), f_{XY})$ be a bivariate random variable with marginal densities f_X and f_Y. Then the conditional density for X given $Y = y$ is defined to be

$$f_{X|y}(x) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

whenever $f_Y(y) > 0$. Ana analogous definition holds for $f_{Y|x}$.

Expectation for Discrete Bivariate Random Variables

1.4.11. Definition. Let $((X, Y), f_{XY})$ be a discrete bivariate random variable and $H \colon \Omega \to \mathbb{R}$ some function. Then the expected value of $H \circ (X, Y)$ is

$$\mathrm{E}[H \circ (X, Y)] = \sum_{(x,y) \in \Omega} H(x, y) \cdot f_{XY}(x, y).$$

provided that the sum (series) on the right converges absolutely. As special cases, we consider $H(x, y) = x$ and $H(x, y) = y$, giving

$$\mathrm{E}[X] = \sum_{(x,y) \in \Omega} x \cdot f_{XY}(x, y), \qquad \mathrm{E}[Y] = \sum_{(x,y) \in \Omega} y \cdot f_{XY}(x, y).$$

Expectation for Continuous Bivariate Random Variables

1.4.14. Definition. Let $((X, Y), f_{XY})$ be a continuous bivariate random variable and $H\colon \mathbb{R}^2 \to \mathbb{R}$ some function. Then the expected value of $H \circ (X, Y)$ is

$$\mathrm{E}[H \circ (X, Y)] = \iint_{\mathbb{R}^2} H(x, y) \cdot f_{XY}(x, y)\, dx\, dy.$$

provided that the integral on the right converges absolutely. As special cases, we consider $H(x, y) = x$ and $H(x, y) = y$, giving

$$\mathrm{E}[X] = \iint_{\mathbb{R}^2} x \cdot f_{XY}(x, y)\, dx\, dy, \qquad \mathrm{E}[Y] = \iint_{\mathbb{R}^2} y \cdot f_{XY}(x, y)\, dx\, dy.$$

Conditional Expectation

1.4.15. Definition. Let $((X, Y), f_{XY})$ be a discrete bivariate random variable. Then we define the conditional expectations

$$\mathrm{E}[Y \mid x] := \sum_y y \cdot f_{Y|x}(y), \qquad \mathrm{E}[X \mid y] := \sum_x x \cdot f_{X|y}(x)$$

where $f_{Y|x}$ and $f_{X|y}$ are the conditional density for Y given x and X given y, respectively.

1.4.16. Definition. Let $((X, Y), f_{XY})$ be a continuous bivariate random variable. Then we define the conditional expectations

$$\mathrm{E}[Y \mid x] := \int_{\mathbb{R}} y \cdot f_{Y|x}(y)\, dy, \qquad \mathrm{E}[X \mid y] := \int_{\mathbb{R}} x \cdot f_{X|y}(x)\, dx$$

where $f_{Y|x}$ and $f_{X|y}$ are the conditional density for Y given x and X given y, respectively.

Conditional Expectation

The discrete case:

$$E[X] = \sum_i E[X|y_i] f_Y(y_i).$$

The continuous case:

$$E[X] = \int_{\{y: f_Y(y) > 0\}} E[X|y] f_Y(y) dy.$$

Why are the two relationships useful?

Sometimes we want to know $E[X]$ and $E[X|y]$ is easier to get than $E[X]$.

Recall the total probability formula (1.1.3) on P53.

Covariance

1.4.18. Definition. Let $((X, Y), f_{XY})$ be a bivariate random variable with means $\mu_X = \mathrm{E}[X]$ and $\mu_Y = \mathrm{E}[Y]$. Then the ***covariance of*** (X, Y) is given by

$$\mathrm{Cov}(X, Y) = \sigma_{XY} = \mathrm{E}[(X - \mu_X)(Y - \mu_Y)].$$

Covariance

$$Var[X+Y] = Var[X] + Var[Y] + 2Cov(X,Y).$$

$$Var[X-Y] = Var[X] + Var[Y] - 2Cov(X,Y).$$

$$Var[XY] = Var[X]Var[Y] + E[X]^2Var[Y] + E[Y]^2Var[X].$$

Covariance

1.4.19. Theorem. Let $((X, Y), f_{XY})$ be a bivariate random variable. If X and Y are independent, then

$$\mathrm{Cov}(X, Y) = 0 \qquad \text{or, equivalently,} \qquad \mathrm{E}[XY] = \mathrm{E}[X]\,\mathrm{E}[Y].$$

Covariance

$$E[XY] = E[X]E[Y] + Cov(X, Y).$$

Note that $E[XY] = E[X]E[Y]$ holds only if X, Y are uncorrelated ($Cov(X, Y) = \rho_{XY} = 0$).
In comparison, $E[X + Y] = E[X] + E[Y]$ holds generally.

Covariance

$$\operatorname{Var} X = \operatorname{Var}(X_1 + \cdots + X_n) = \operatorname{Var} X_1 + \cdots + \operatorname{Var} X_n + 2 \sum_{i<j} \operatorname{Cov}(X_i, X_j).$$

For Your reference

$Var[X_i] = \frac{r}{N}\left(1 - \frac{r}{N}\right)$ **Proof:**

$$Var[X_i] = E[X_i^2] - E[X_i]^2 = \frac{r}{N} - \left(\frac{r}{N}\right)^2 = \frac{r}{N}\left(1 - \frac{r}{N}\right).$$

$Cov(X_i, X_j) = -\frac{1}{N} \cdot \frac{r(N-r)}{N(N-1)}$ **Proof:**

For $i \neq j$,

$$\begin{aligned} Cov(X_i, X_j) &= E[X_iX_j] - E[X_i]E[X_j] \\ &= \frac{r}{N}\frac{r-1}{N-1} - \left(\frac{r}{N}\right)^2 \\ &= \frac{r(r-1)N - r^2(N-1)}{N^2(N-1)} \\ &= -\frac{1}{N} \cdot \frac{r(N-r)}{N(N-1)}. \end{aligned}$$

For Your reference

$Var[X] = n\frac{r}{N}\frac{N-r}{N}\frac{N-n}{N-1}$ **Proof:**

$$
\begin{aligned}
Var[X] &= n\frac{r}{N}\left(1-\frac{r}{N}\right) + 2\frac{n(n-1)}{2}\left(-\frac{1}{N}\right)\cdot\frac{r(N-r)}{N(N-1)} \\
&= \frac{nr(N-r)(N-1) - n(n-1)r(N-r)}{N^2(N-1)} \\
&= n\frac{r}{N}\frac{N-r}{N}\frac{N-n}{N-1}.
\end{aligned}
$$

For Your reference

$Var[W] = 1$ **Proof:**

$$Var[W] = E\left[\left(\frac{X-\mu_X}{\sigma_X}\right)^2\right] - E[W]^2 = \frac{E[X^2] - 2E[X]^2 + E[X]^2}{\sigma_X^2} = 1.$$

For Your reference

$E[WZ] = \frac{Cov(X,Y)}{\sqrt{Var[X]Var[Y]}}$ **Proof:**

$$
\begin{aligned}
E[WZ] &= E\left[\frac{X-\mu_X}{\sigma_X}\frac{Y-\mu_Y}{\sigma_Y}\right] \\
&= \frac{E[XY]-2E[X]E[Y]+E[X]E[Y]}{\sigma_X\sigma_Y} \\
&= \frac{Cov(X,Y)}{\sqrt{Var[X]Var[Y]}}.
\end{aligned}
$$

Important!

If X, Y are independent, X, Y are uncorrelated.

If X, Y are uncorrelated, X, Y are not necessarily independent.

Find a counterexample in your assignment!

The Cauchy-Schwartz inequality

1.4.21. Theorem. Let $((X, Y), f_{XY})$ be a bivariate random variable with correlation coefficient ρ_{XY}.

(i) $-1 \leq \rho_{XY} \leq 1$.

(ii) $|\rho_{XY}| = 1$ if and only if there exist numbers $\beta_0, \beta_1 \in \mathbb{R}$, $\beta_1 \neq 0$, such that

$$Y = \beta_0 + \beta_1 X$$

almost surely.

The Bivariate Normal Distribution

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\varrho^2}}\, e^{-\frac{1}{2(1-\varrho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\varrho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]} \tag{1.4.2}$$

where $-1 < \varrho < 1$.

Transformation of Variables

1.4.22. Theorem. Let $((X, Y), f_{XY})$ be a continuous bivariate random variable and let $H\colon \mathbb{R}^2 \to \mathbb{R}^2$ be a differentiable bijective map with inverse H^{-1}. Then $(U, V) = H \circ (X, Y)$ is a continuous bivariate random variable with density

$$f_{UV}(u, v) = f_{XY} \circ H^{-1}(u, v) \cdot |\det DH^{-1}(u, v)|,$$

where DH^{-1} is the Jacobian of H^{-1}.

Transformation of Variables

1.4.23. Theorem. Let $((X, Y), f_{XY})$ be a continuous bivariate random variable. Let $U = X/Y$. Then the density f_U of U is given by

$$f_U(u) = \int_{-\infty}^{\infty} f_{XY}(uv, v) \cdot |v| \, dv.$$

Transformation of Variables

The case of $U = X + Y$ will be proved in the assignment and this result is very useful.

For the case of $U = XY$,

$$f_U(u) = \int_{-\infty}^{\infty} |v|^{-1} f_{XY}(v^{-1}u, v)dv.$$

You can prove it by yourself.

Ex.1 (Marginal Density Functions & Joint Density Functions)

X, Y are two discrete random variables. Suppose that X can be 0 or 1 and Y can be 0 or 1.

$$f_{XY}(0,0) = 1/4 + \varepsilon,$$
$$f_{XY}(0,1) = 1/4 - \varepsilon,$$
$$f_{XY}(1,0) = 1/4 - \varepsilon,$$
$$f_{XY}(1,1) = 1/4 + \varepsilon.$$

We have $f_X(0) = f_X(1) = f_Y(0) = f_Y(1) = 1/2$. We can see that two marginal density functions have nothing to do with ε. Therefore, for discrete random variables, marginal density functions cannot determine the joint density function.

What about the case for continuous random variables?

This conclusion also holds.

Refer to the formula (1.4.2) on P231. If we fix $\mu_X, \mu_Y, \sigma_X, \sigma_Y$, then two marginal density functions are fixed. And if we change ρ, the joint density function varies.

On the other hand, marginal density functions are completely determined by joint density functions.

Ex.2 (Rayleigh Distribution)

Let X, Y be i.i.d random variables and $X, Y \sim N(0,1)$. $Z = \sqrt{X^2 + Y^2}$. Find $f_Z(z)$.

Solution:

$f_Z(z) = 0$ for any $z \leq 0$.

For $z > 0$,

$$F_Z(z) = \iint\limits_{\{(x,y):x^2+y^2\leq z^2\}} \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)} dxdy.$$

Let $r^2 = x^2 + y^2$.

$$F_Z(z) = \int_0^{2\pi} \int_0^z \frac{1}{2\pi} e^{-\frac{1}{2}r^2} r dr d\theta = \int_0^z e^{-\frac{1}{2}r^2} r dr.$$

Therefore,

$$f_Z(z) = \begin{cases} 0, & z \leq 0, \\ ze^{-\frac{1}{2}z^2}, & z > 0. \end{cases}$$

This is a Rayleigh distribution. The generalized Rayleigh distribution also has a parameter σ (here we let $\sigma = 1$ for simplification).

Verify by yourself that $E[Z] = \sqrt{\frac{\pi}{2}}\sigma$ and $Var[Z] = \frac{4-\pi}{2}\sigma^2$. (Hint: integration by brute force)

I strongly recommend you to do this exercise. If you can do it all by yourself, you will not be afraid of any integration in Ve401.

One last comment is that this exercise is essentially different from P233 1.4.22! This exercise is 2 dimensions to 1 dimension. 1.4.22 is 2 dimensions to 2 dimensions. However, you can also solve 2 dimensions to 1 dimension questions by 1.4.22 like what we have done in the proof of 1.4.23. My recommendation is to first look at dimensions and then choose a best method.

Final Tips

- Lecture slides
- Homework
  1. methods
  2. correct answers
  3. theorems

- Calculator allowed
- No cheating paper