VE401 Probabilistic Methods in Eng.
RC 3

CHEN Xiwen

UM-SJTU Joint Institute

March 22, 2020

Table of contents

Normal Distribution

Definition. A continuous random variable (X, f_{μ,σ^2}) has the ***normal distribution*** with mean $\mu \in \mathbb{R}$ and variance $\sigma^2, \sigma > 0$ if the probability density function is given by

$$f_{\mu,\sigma^2} = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \qquad x \in \mathbb{R}.$$

Normal Distribution

Mean, variance and M.G.F.

- Mean.

$$E[X] = \mu.$$

- Variance.

$$\mathrm{Var}[X] = \sigma^2.$$

- M.G.F.

$$m_X : \mathbb{R} \to \mathbb{R}, \qquad m_X(t) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right).$$

Normal Distribution

Verifying M.G.F.

$$
\begin{aligned}
m_X(t) = \mathrm{E}\left[e^{tX}\right] &= \int_{-\infty}^{\infty} \frac{e^{tx}}{\sqrt{2\pi}\sigma} e^{-((x-\mu)/\sigma)^2/2} \mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{\mu t + \sigma^2 t^2/2} \cdot e^{-\frac{(x-(\mu+\sigma^2 t))^2}{2\sigma^2}} \mathrm{d}x \\
&= e^{\mu t + \sigma^2 t^2/2} \underbrace{\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{(x-(\mu+\sigma^2 t))^2}{2\sigma^2}} \mathrm{d}x}_{=1} \\
&= e^{\mu t + \sigma^2 t^2/2}.
\end{aligned}
$$

Normal Distribution

Some takeaway from this proof.

- To verify that

$$I := \int_{-\infty}^{\infty} e^{-\frac{(x-b)^2}{a^2}} \mathrm{d}x = a\sqrt{\pi},$$

we use

$$I^2 = \left(\int_{-\infty}^{\infty} e^{-\frac{(x-a)^2}{b^2}} \mathrm{d}x\right)^2 = \int_{-\infty}^{\infty} e^{-\frac{(x-a)^2}{b^2}} \cdot e^{-\frac{(y-a)^2}{b^2}} \mathrm{d}x\mathrm{d}y.$$

Using parametrization $x = ar\cos\theta + b, y = ar\sin\theta + b$, we have

$$\begin{aligned} I^2 &= \int_0^{\infty} \int_0^{2\pi} e^{-r^2} \cdot a^2 r \mathrm{d}\theta \mathrm{d}r \\ &= a^2\pi \int_0^{\infty} 2re^{-r^2} \mathrm{d}r = -a^2\pi e^{-r^2}\Big|_0^{\infty} = a^2\pi. \end{aligned}$$

Normal Distribution

Some takeaway from this proof.

- Useful results from normalizing constant of distributions.
  (i). Normal.

$$\int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \mathrm{d}x = \sqrt{2\pi}\sigma.$$

  (ii). Gamma.

$$\int_{0}^{\infty} x^{\alpha-1} e^{-\beta x} \mathrm{d}x = \frac{\Gamma(\alpha)}{\beta^{\alpha}}.$$

Transformation of Random Variables

- Discrete random variables. Let X be a discrete random variable with probability density function f_X, the the probability density function f_Y for $Y = \varphi(X)$ is given by

$$f_Y(y) = \sum_{x \in \varphi^{-1}(y)} f_X(x), \qquad \text{for } y \in \text{ran } \varphi,$$

and 0 otherwise.

Example 1. Let X be a uniform random variable on $\{-n, -n+1, \ldots, n-1, n\}$. Then $Y = |X|$ has probability density function

$$f_Y(y) = \begin{cases} \dfrac{1}{2n+1} & x = 0, \\ \dfrac{2}{2n+1} & x \neq 0. \end{cases}$$

Transformation of Random Variables

- Continuous random variables. Let X be a continuous random variable with density f_X. Let $Y = \varphi \circ X$, where $\varphi : \mathbb{R} \to \mathbb{R}$ is strictly monotonic and differentiable. The density for Y is then given by

$$f_Y(y) = f_X(\varphi^{-1}(y)) \cdot \left| \frac{\mathrm{d}\varphi^{-1}(y)}{\mathrm{d}y} \right|, \qquad \text{for } y \in \operatorname{ran} \varphi$$

and

$$f_Y(y) = 0, \qquad \text{for } y \notin \operatorname{ran} \varphi.$$

Transformation of Random Variables

- Continuous random variables. Let X be a continuous random variable with density f_X. Let $Y = \varphi \circ X$, where $\varphi : \mathbb{R} \to \mathbb{R}$ is strictly monotonic and differentiable. The density for Y is then given by

$$f_Y(y) = f_X(\varphi^{-1}(y)) \cdot \left| \frac{\mathrm{d}\varphi^{-1}(y)}{\mathrm{d}y} \right|, \qquad \text{for } y \in \text{ran } \varphi$$

and

$$f_Y(y) = 0, \qquad \text{for } y \notin \text{ran } \varphi.$$

For multivariate random variables, $\mathbf{Y} = \varphi \circ \mathbf{X}$, we have

$$f_{\mathbf{Y}}(y) = f_{\mathbf{X}} \circ \varphi^{-1}(y) \cdot |\det \; D\varphi^{-1}(y)|,$$

where $D\varphi^{-1}$ is the Jacobian of φ^{-1}.

From RC2 Part 1: Connections of Discrete Distributions

- Bernoulli $\rightarrow$ Binomial. $X_1, \ldots, X_n$ are independent random variables,

$$X_i \sim \text{Bernoulli}(p) \quad \Rightarrow \quad X = X_1 + \cdots + X_n \sim \text{B}(n, p).$$

- Binomial $\rightarrow$ Binomial. $X_1, \ldots, X_k$ are independent random variables,

$$X_i \sim \text{B}(n_i, p) \quad \Rightarrow \quad X = X_1 + \cdots + X_k \sim \text{B}(n, p),$$

  where $n = n_1 + \cdots + n_k$.

- Geometric $\rightarrow$ Negative binomial. $X_1, \ldots, X_r$ are independent random variables,

$$X_i \sim \text{Geom}(p) \quad \Rightarrow \quad X = X_1 + \cdots + X_r \sim \text{NB}(r, p).$$

From RC2 Part 1: Connections of Discrete Distributions

- Negative binomial $\to$ Negative binomial. $X_1, \ldots, X_n$ are independent random variables,

$$X_i \sim \text{NB}(r_i, p) \quad \Rightarrow \quad X = X_1 + \cdots + X_n \sim \text{NB}(r, p),$$

where $r = r_1 + \cdots + r_n$.

- Poisson $\to$ Poisson. $X_1, \ldots, X_n$ are independent random variables,

$$X_i \sim \text{Poisson}(k_i) \quad \Rightarrow \quad X = X_1 + \cdots + X_n \sim \text{Poisson}(k),$$

where $k = k_1 + \cdots + k_n$.

From RC2 Part 1: Connections of Discrete Distributions

- Negative binomial $\rightarrow$ Negative binomial. $X_1, \ldots, X_n$ are independent random variables,

$$X_i \sim \text{NB}(r_i, p) \quad \Rightarrow \quad X = X_1 + \cdots + X_n \sim \text{NB}(r, p),$$

where $r = r_1 + \cdots + r_n$.

- Poisson $\rightarrow$ Poisson. $X_1, \ldots, X_n$ are independent random variables,

$$X_i \sim \text{Poisson}(k_i) \quad \Rightarrow \quad X = X_1 + \cdots + X_n \sim \text{Poisson}(k),$$

where $k = k_1 + \cdots + k_n$.

Digression. A Second Look into Connections of Distributions — `s3.pdf`.

Sum of Normal Distributions

Theorem. If the random variables $X_1, \ldots, X_k$ are independent and if X_i has the normal distribution with mean μ_i and variances σ_i^2, where $i = 1, \ldots, k$, then the sum

$$X = X_1 + \cdots + X_k$$

follows the normal distribution with

$$\mu = \mu_1 + \cdots + \mu_k, \qquad \sigma^2 = \sigma_1^2 + \cdots + \sigma_k^2.$$

Sum of Normal Distributions

Theorem. If the random variables $X_1, \dots, X_k$ are independent and if X_i has the normal distribution with mean μ_i and variances σ_i^2, where $i = 1, \dots, k$, then the sum

$$X = X_1 + \cdots + X_k$$

follows the normal distribution with

$$\mu = \mu_1 + \cdots + \mu_k, \qquad \sigma^2 = \sigma_1^2 + \cdots + \sigma_k^2.$$

Proof (sketch). Using M.G.F., we have

$$\begin{aligned} m_X(t) &= \prod_{i=1}^{k} m_{X_i}(t) = \prod_{i=1}^{k} \exp\left(\mu_i t + \frac{1}{2}\sigma_i^2 t^2\right) \\ &= \exp\left[\left(\sum_{i=1}^{k} \mu_i\right) t + \frac{1}{2}\left(\sum_{i=1}^{k} \sigma_i^2\right) t^2\right], \qquad t \in \mathbb{R}. \end{aligned}$$

Quotient of Normal Distributions

Theorem. Suppose that random variables X and Y are independent and that each has the standard normal distribution. Then $U = X/Y$ has the *Cauchy distribution* with probability density function given by

$$f_U(u) = \frac{1}{\pi(1+u^2)}, \qquad u \in \mathbb{R}.$$

Quotient of Normal Distributions

Theorem. Suppose that random variables X and Y are independent and that each has the standard normal distribution. Then $U = X/Y$ has the *Cauchy distribution* with probability density function given by

$$f_U(u) = \frac{1}{\pi(1+u^2)}, \qquad u \in \mathbb{R}.$$

Proof (sketch). Let $V = Y$, excluding $Y = 0$, the transformation from (X, Y) to (U, V) is one-to-one. Then $X = UV$, $Y = V$ and

$$J = \det \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix} = v.$$

Quotient of Normal Distributions

Theorem. Suppose that random variables X and Y are independent and that each has the standard normal distribution. Then $U = X/Y$ has the *Cauchy distribution* with probability density function given by

$$f_U(u) = \frac{1}{\pi(1+u^2)}, \qquad u \in \mathbb{R}.$$

Proof (sketch, continued). Then the joint density function is given by

$$f_{UV}(u,v) = f_{XY}(uv,v)|v| = \frac{|v|}{2\pi}\exp\left(-\frac{1}{2}(u^2+1)v^2\right).$$

Then the marginal of U is calculated as

$$f_U(u) = \int_{-\infty}^{\infty} f_{UV}(u,v)\mathrm{d}v = \frac{1}{\pi(u^2+1)}, \quad u \in \mathbb{R}.$$

Standardizing Normal Distribution

Suppose $X \sim \text{Normal}(\mu, \sigma^2)$. Then

$$Z = \frac{X - \mu}{\sigma} \sim \text{Normal}(0, 1),$$

where the normal distribution with mean μ and variance σ^2 is the ***standard normal distribution***. Furthermore, the cumulative distribution function of X is given by

$$F(x) = \Phi\left(\frac{x - \mu}{\sigma}\right), \quad F^{-1}(p) = \mu + \sigma\Phi^{-1}(p),$$

where Φ is the cumulative distribution function for the standard normal distribution function.

Common Applications of Normal Distribution

Suppose a random variable X follows normal distribution $N(\mu, \sigma)$, where μ and σ are known. At current stage, applications usually include the following.

1. Given some value x_0, find the probability of $P[X \leq x_0]$ or $P[X \geq x_0]$.
   (a). Standardize X as $Z = (X - \mu)/\sigma$, find z_0.
   (b). Find $P[X \leq x_0] = P[Z \leq z_0], P[X \geq x_0] = 1 - P[Z \geq z_0]$.
2. Given some probability p, find the corresponding x_0 such that $P[X \leq x_0] = p$ or $P[X \geq x_0] = p$.
   (a). Find z_0 from table such that $P[Z \leq z_0] = p$ or $P[Z \leq z_0] = 1 - p$.
   (b). Calculate $x_0 = \sigma z_0 + \mu$.
3. "Three-sigma" rule.

$$P[-3\sigma < X - \mu < 2\sigma] = 0.997.$$

The Chebyshev's Inequality

Theorem. Let X be a random variable, then for $k \in \mathbb{N} \setminus \{0\}$ and $c > 0$,

$$P[|X| \geq c] \leq \frac{\mathrm{E}[|X|^k]}{c^k}.$$

As another version of this inequality, suppose X has mean μ and standard deviation σ, and let $m > 0$,

$$P[|X - \mu| \geq m\sigma] \leq \frac{1}{m^2},$$

or equivalently,

$$P[-m\sigma < X - \mu < m\sigma] \geq 1 - \frac{1}{m^2}.$$

Note. This yields another (looser) version of $\sigma, 2\sigma, 3\sigma$ rule for normal distribution.

Application of Chebyshev's Inequality

Weak Law of Large Numbers. Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with mean μ and variance σ^2. Then for any $\varepsilon > 0$,

$$P\left[\left|\frac{X_1 + \ldots + X_n}{n} - \mu\right| \geq \varepsilon\right] \xrightarrow{n\to\infty} 0.$$

Application of Chebyshev's Inequality

Weak Law of Large Numbers. Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with mean μ and variance σ^2. Then for any $\varepsilon > 0$,

$$P\left[\left|\frac{X_1 + \ldots + X_n}{n} - \mu\right| \geq \varepsilon\right] \xrightarrow{n\to\infty} 0.$$

Law of Large Numbers. Let A be a random outcome (random event) of an experiment that can be repeated without the outcome influencing subsequent repetitions. Then the probability $P[A]$ of this event occurring may be approximated by

$$P[A] \approx \frac{\text{number of times } A \text{ occurs}}{\text{number of times experiment is perfomred}}.$$

Note. Approximate mean $\mu = p = P[A]$ of Bernoulli distribution.

Application of Chebyshev's Inequality

Weak Law of Large Numbers. Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with mean μ and variance σ^2. Then for any $\varepsilon > 0$,

$$P\left[\left|\frac{X_1 + \ldots + X_n}{n} - \mu\right| \geq \varepsilon\right] \xrightarrow{n\to\infty} 0.$$

Proof. Using properties of expectation and variance,

$$\mathrm{E}\left[\frac{X_1 + \cdots + X_n}{n} - \mu\right] = \frac{\mathrm{E}[X_1] + \cdots + \mathrm{E}[X_n]}{n} - \mathrm{E}[\mu] = 0,$$

$$\mathrm{Var}\left[\frac{X_1 + \cdots + X_n}{n} - \mu\right] = \frac{\mathrm{Var}[X_1] + \cdots + \mathrm{Var}[X_n]}{n^2} + \mathrm{Var}[\mu] = \frac{\sigma^2}{n},$$

$$\Rightarrow \quad \mathrm{E}\left[\left(\frac{X_1 + \cdots + X_n}{n} - \mu\right)^2\right] = \frac{\sigma^2}{n}.$$

Application of Chebyshev's Inequality

Weak Law of Large Numbers. Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with mean μ and variance σ^2. Then for any $\varepsilon > 0$,

$$P\left[\left|\frac{X_1 + \ldots + X_n}{n} - \mu\right| \geq \varepsilon\right] \xrightarrow{n\to\infty} 0.$$

Proof (continued). Applying the Chebyshev's inequality with $k = 2$ to

$$X = \frac{X_1 + \cdots + X_n}{n} - \mu,$$

we have

$$P\left[\left|\frac{X_1 + \ldots + X_n}{n} - \mu\right| \geq \varepsilon\right] \leq \frac{\sigma^2}{n\varepsilon^2} \xrightarrow{n\to\infty} 0.$$

Normal Approximation of Binomial Distribution

Suppose S_n is the number of successes in a sequence of n i.i.d. Bernoulli trials with probability of success $0 < p < 1$.

- It satisfies that

$$\lim_{n\to\infty} P\left[a < \frac{X - np}{\sqrt{np(1-p)}} \le b\right] = \frac{1}{2\pi}\int_a^b e^{-x^2/2} \mathrm{d}x.$$

- For $y = 0, \ldots, n$,

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

  where we require that

$$np > 5 \quad \text{if } p \le \frac{1}{2} \qquad \text{or} \qquad n(1-p) > 5 \quad \text{if } p > \frac{1}{2}.$$

Discrete Multivariate Random Variables

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

$$\mathbf{X} : S \to \Omega$$

together with a function $f_{\mathbf{X}} : \Omega \to \mathbb{R}$ with the properties that

(i). $f_{\mathbf{X}}(x) \geq 0$ for all $x = (x_1, \ldots, x_n) \in \Omega$ and

(ii). $\displaystyle\sum_{x \in \Omega} f_{\mathbf{X}}(x) = 1$,

where $f_{\mathbf{X}}$ is the ***joint density function*** of the random variable $\mathbf{X}$.

Discrete Multivariate Random Variables

Definition.

- ***Marginal density*** f_{X_k} for $X_k, k = 1, \ldots, n$:

$$f_{X_k}(x_k) = \sum_{x_1,\ldots,x_{k-1},x_{k+1},\ldots,x_n} f_{\mathbf{X}}(x_1, \ldots, x_n).$$

- ***Independent*** multivariate random variables:

$$f_{\mathbf{X}}(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n).$$

- ***Conditional density*** of X_1 conditioned on X_2:

$$f_{X_1|X_2}(x_1) := \frac{f_{X_1 X_2}(x_1, x_2)}{f_{X_2}(x_2)} \qquad \text{whenever } f_{X_2}(x_2) > 0.$$

Continuous Multivariate Random Variables

Definition. Let S be a sample space. A ***continuous multivariate random variable*** is a map

$$\mathbf{X} : S \to \mathbb{R}^n$$

together with a function $f_{\mathbf{X}} : \mathbb{R}^n \to \mathbb{R}$ with the properties that

(i). $f_{\mathbf{X}}(x) \geq 0$ for all $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and

(ii). $\displaystyle\int_{\mathbb{R}^n} f_{\mathbf{X}}(x) = 1$,

where $f_{\mathbf{X}}$ is the ***joint density function*** of the random variable $\mathbf{X}$.

Continuous Multivariate Random Variables

Definition.

- ***Marginal density*** f_{X_k} for $X_k, k = 1, \dots, n$:

$$f_{X_k}(x_k) = \int_{\mathbb{R}^{n-1}} f_{\mathbf{X}}(x_1, \dots, x_n)\mathrm{d}x_1 \dots \mathrm{d}x_{k-1}\mathrm{d}x_{k+1} \dots \mathrm{d}x_n.$$

- ***Independent*** multivariate random variables:

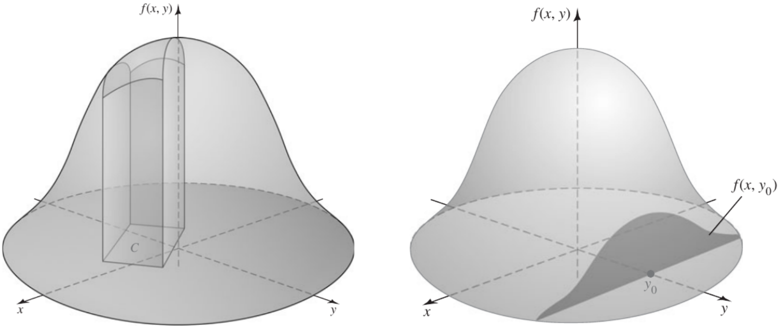
$$f_{\mathbf{X}}(x_1, \dots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n).$$

- ***Conditional density*** of X_1 conditioned on X_2:

$$f_{X_1|X_2}(x_1) := \frac{f_{X_1 X_2}(x_1, x_2)}{f_{X_2}(x_2)} \qquad \text{whenever } f_{X_2}(x_2) > 0.$$

Continuous Multivariate Random Variables

Visualization. Joint probability density function $f_{XY}(x,y)$ (left) and conditional density function $f_{X|Y}(x|y_0)$ (right).



Continuous Multivariate Random Variables

Q. How to determine the joint probability density function and cumulative distribution function of a single variable from a joint cumulative distribution function?

Continuous Multivariate Random Variables

Q. How to determine the joint probability density function and cumulative distribution function of a single variable from a joint cumulative distribution function?

C.D.F. For continuous random variables $X_1, \ldots, X_n$, the joint cumulative distribution function is then given by

$$P[X_1 \leq a_1, \ldots, X_n \leq a_n] = \int_{-\infty}^{a_1} \cdots \int_{-\infty}^{a_n} f_{\mathbf{X}}(x) \mathrm{d}x_1 \ldots \mathrm{d}x_n.$$

Continuous Multivariate Random Variables

Example 2. Suppose X and Y are random variables that take values in the intervals $0 \leq X \leq 2$ and $0 \leq Y \leq 2$. Suppose the joint cumulative distribution function for $x \in [0, 2], y \in [0, 2]$ is given by

$$F(x, y) = \frac{1}{16} xy(x + y).$$

What are the joint density function and cumulative distribution of X?

Continuous Multivariate Random Variables

Example 2. Suppose X and Y are random variables that take values in the intervals $0 \le X \le 2$ and $0 \le Y \le 2$. Suppose the joint cumulative distribution function for $x \in [0,2], y \in [0,2]$ is given by

$$F(x,y) = \frac{1}{16}xy(x+y).$$

What are the joint density function and cumulative distribution of X?

Solution (i). For $x \in [0,2], y \in [0,2]$,

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y} = \frac{1}{8}(x+y),$$

and thus

$$f_{XY}(x,y) = \begin{cases} \frac{1}{8}(x+y) & 0 \le x \le 2, 0 \le y \le 2 \\ 0 & \text{otherwise.} \end{cases}$$

Continuous Multivariate Random Variables

Example 2. Suppose X and Y are random variables that take values in the intervals $0 \leq X \leq 2$ and $0 \leq Y \leq 2$. Suppose the joint cumulative distribution function for $x \in [0,2], y \in [0,2]$ is given by

$$F(x,y) = \frac{1}{16}xy(x+y).$$

What are the joint density function and cumulative distribution of X?

Solution (ii). Since for $y > 2$, $F(x,y) = F(x,2)$, then by letting $y \to \infty$, we obtain

$$F_X(x) = \begin{cases} 0 & x < 0, \\ \frac{1}{8}x(x+2) & 0 \leq x \leq 2, \\ 1 & x > 2. \end{cases}$$

Expectation

- <u>Discrete</u>.

$$\mathrm{E}[X_k] = \sum_{x_k} x_k f_{X_k}(x_k) = \sum_{x \in \Omega} x_k f_{\mathbf{X}}(x),$$

  and for continuous function $\varphi : \mathbb{R}^n \to \mathbb{R}$,

$$\mathrm{E}[\varphi \circ \mathbf{X}] = \sum_{x \in \Omega} \varphi(x) f_{\mathbf{X}}(x).$$

- <u>Continuous</u>.

$$\mathrm{E}[X_k] = \int_{\mathbb{R}} x_k f_{X_k}(x_k) \mathrm{d}x_k = \int_{\mathbb{R}^n} x_k f_{\mathbf{X}}(x) \mathrm{d}x,$$

  and for continuous function $\varphi : \mathbb{R}^n \to \mathbb{R}$,

$$\mathrm{E}[\varphi \circ \mathbf{X}] = \int_{\mathbb{R}^n} \varphi(x) f_{\mathbf{X}}(x) \mathrm{d}x.$$

Covariance and Covariance Matrix

Definition. For a multivariate random variable $\mathbf{X}$, the ***covariance matrix*** is given by

$$\mathrm{Var}[\mathbf{X}] = \begin{pmatrix} \mathrm{Var}[X_1] & \mathrm{Cov}[X_1, X_2] & \cdots & \mathrm{Cov}[X_1, X_n] \\ \mathrm{Cov}[X_1, X_2] & \mathrm{Var}[X_2] & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathrm{Cov}[X_{n-1}, X_n] \\ \mathrm{Cov}[X_1, X_n] & \cdots & \mathrm{Cov}[X_{n-1}, X_n] & \mathrm{Var}[X_n] \end{pmatrix},$$

where the ***covariance*** of (X_i, X_j) is given by

$$\mathrm{Cov}[X_i, X_j] = \mathrm{E}[(X_i - \mu_{X_i})(X_j - \mu_{X_j})] = \mathrm{E}[X_i X_j] - \mathrm{E}[X_i]\mathrm{E}[X_j],$$

and

$$\mathrm{Var}[\mathbf{CX}] = \mathbf{C}\mathrm{Var}[\mathbf{X}]\mathbf{C}^T, \qquad \mathbf{C} \in \mathrm{Mat}(n \times n; \mathbb{R}).$$

Covariance and Independence

Let $X, X_1, \ldots, X_n$ and Y be random variables.

- X and Y are independent $\Rightarrow \mathrm{Cov}[X, Y] = 0$, while the converse is ***not*** true.
- $\mathrm{Var}[X+Y] = \mathrm{Var}[X]+\mathrm{Var}[Y]+2\mathrm{Cov}[X, Y]$, and more generally,

$$\mathrm{Var}[X_1 + \cdots + X_n] = \mathrm{Var}[X_1] + \cdots + \mathrm{Var}[X_n] + 2\sum_{i<j} \mathrm{Cov}[X_i, X_j],$$

if $\mathrm{Var}[X_i] < \infty$ for $i = 1, \ldots, n$.

Covariance and Independence

Example 3. Suppose the random variable X can take only three values -1, 0, and 1, and each of these values has the same probability. Also, let random variable Y satisfy $Y = X^2$. Then X and Y are apparently dependent, while

$$E[XY] = E[X^3] = E[X] = 0,$$

and thus

$$\text{Cov}[X, Y] = E[XY] - E[X]E[Y] = 0.$$

Pearson Correlation Coefficient

Definition. The ***Pearson coefficient of correlation*** of random variables X and Y is given by

$$\rho_{XY} := \frac{\mathrm{Cov}[X, Y]}{\sqrt{\mathrm{Var}[X]\mathrm{Var}[Y]}}.$$

Note. Instead of independence, the correlation coefficient actually measures the the extent to which X and Y are linearly dependent, which is not the only way of being dependent.

Properties.

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

(ii). $|\rho_{XY}| = 1$ iff there exist $\beta_0, \beta_1 \in \mathbb{R}$ such that

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

The Fisher Transformation

Definition. Let $\tilde{X}$ and $\tilde{Y}$ be standardized random variables of X and Y, then the ***Fisher transformation*** of ρ_{XY} is given by

$$\ln\left(\sqrt{\frac{\mathrm{Var}[\tilde{X}+\tilde{Y}]}{\mathrm{Var}[\tilde{X}-\tilde{Y}]}}\right) = \frac{1}{2}\ln\left(\frac{1+\rho_{XY}}{1-\rho_{XY}}\right) = \mathrm{Arctanh}(\rho_{XY}) \in \mathbb{R}.$$

We say that X and Y are

- ***positively correlated*** if $\rho_{XY} > 0$, and
- ***negatively correlated*** if $\rho_{XY} < 0$.

The Hypergeometirc Distribution

Definition. A random variable (X, f_X) with parameters $N, n, r \in \mathbb{N}\backslash\{0\}$ where $r, n \leq N$ and $n < \min\{r, N-r\}$ has a ***hypergeometric distribution*** if the density function is given by

$$f_X(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}.$$

Interpretation.

- $f_X(x)$ is the probability of getting x balls in drawing n balls from a box containing N balls, where r of them are red.
- This can be formulated as obtaining x successes in n identical but ***not*** independent Bernoulli trials, each with probability of success $\frac{r}{N}$.

The Hypergeometirc Distribution

- Expectation.

$$\mathrm{E}[X] = \mathrm{E}[X_1 + \cdots + X_n] = n\frac{r}{N}.$$

- Variance.

$$\begin{aligned}\mathrm{Var}[X] &= \mathrm{Var}[X_1 + \cdots + X_n] \\ &= \mathrm{Var}[X_1] + \cdots + \mathrm{Var}[X_n] + 2\sum_{i<j} \mathrm{Cov}[X_i, X_j] \\ &= n\frac{r}{N}\frac{N-r}{N}\frac{N-n}{N-1}.\end{aligned}$$

The binomial distribution may be used to approximate the hypergeometric distribution if n/N is small.

The Hypergeometirc Distribution

Calculation of mean and variance. Transform to Bernoulli trials $(X_1, \ldots, X_n)$.

- The Bernoulli trials are identical with $p_k = \dfrac{r}{N}$, i.e.,

$$
\begin{aligned}
P[X_1 = 1] &= \frac{r}{N}, \\
P[X_2 = 1] &= P[X_2 = 1 | X_1 = 1] P[X_1 = 1] + \\
&\qquad + P[X_2 = 1 | X_1 = 0] P[X_1 = 0] \\
&= \frac{r-1}{N-1} \cdot \frac{r}{N} + \frac{r}{N-1} \frac{N-r}{N} \\
&= \frac{r}{N},
\end{aligned}
$$

  and so on.

The Hypergeometirc Distribution

Calculation of mean and variance. Transform to Bernoulli trials $(X_1, \ldots, X_n)$.

- $\mathrm{E}[X_k] = p_k = \dfrac{r}{N}, \mathrm{Var}[X_k] = p_k(1 - p_k)$.
- For variance,

$$\mathrm{Var}[X] = \sum_{k=1}^{n} \mathrm{Var}[X_k] + 2\sum_{i<j} \mathrm{Cov}[X_i, X_j],$$

where

$$\mathrm{Cov}[X_i, X_j] = \mathrm{E}[X_i X_j] - \mathrm{E}[X_i]\mathrm{E}[X_j],$$
$$\mathrm{E}[X_i, X_j] = p_{ij} = \frac{r}{N} \cdot \frac{r-1}{N-1}, \qquad i \neq j.$$

Closeness of Binomial and Hypergeometirc Distributions

Theorem. Suppose Y has a binomial distribution with parameters $n \in \mathbb{N} \setminus \{0\}$ and $p, 0 < p < 1$. Let $\{X_k\}$ be a sequence of hypergeometric random variables with parameters N_k, n, r_k such that

$$\lim_{k\to\infty} N_k = \infty, \quad \lim_{k\to\infty} r_k = \infty, \quad \lim_{k\to\infty} \frac{r_k}{N_k} = p.$$

Then for each fixed n and each $x = 0, \ldots, n$,

$$\lim_{k\to\infty} \frac{P[Y = x]}{P[X_k = x]} = 1.$$

A proof of this theorem can be found in `s3.pdf`.

The Hypergeometirc Distribution

Example 4. Consider a group of T persons, and let $a_1, \ldots, a_T$ be the heights of these T persons. Suppose that n persons are selected from this group at random without replacement, and let X denote the sum of heights of these n persons. Determine the mean and variance of X.

The Hypergeometirc Distribution

Example 4. Consider a group of T persons, and let $a_1, \ldots, a_T$ be the heights of these T persons. Suppose that n persons are selected from this group at random without replacement, and let X denote the sum of heights of these n persons. Determine the mean and variance of X.

Solution. Let X_i be the height of the i-th person selected. Then $X = X_1 + \cdots + X_n$. Since X_i is equally likely to have any one of the T values,

$$\mathrm{E}[X_i] = \frac{1}{T}\sum_{i=1}^{T} a_i = \mu, \quad \mathrm{Var}[X_i] = \frac{1}{T}\sum_{i=1}^{T}(a_i - \mu)^2 = \sigma^2.$$

Therefore, $E[X] = n\mu$, and

$$\mathrm{Var}[X] = \sum_{i=1}^{n} \mathrm{Var}[X_i] + 2\sum_{i<j} \mathrm{Cov}[X_i, X_j].$$

How to calculate $\mathrm{Cov}[X_i, X_j]$?

The Hypergeometirc Distribution

Example 4. Consider a group of T persons, and let $a_1, \ldots, a_T$ be the heights of these T persons. Suppose that n persons are selected from this group at random without replacement, and let X denote the sum of heights of these n persons. Determine the mean and variance of X.

Solution (approach 1). Knowing that

$$\mathrm{E}[X_i X_j] = \frac{2}{T(T-1)} \sum_{i<j} a_i a_j,$$

and

$$\begin{aligned}
\mathrm{Var}[X_i] &= \frac{1}{T}\sum_{i=1}^{T}(a_i-\mu)^2 = \frac{1}{T}\sum_{i=1}^{T}(a_i^2 - 2\mu a_i + \mu^2) \\
&= \frac{1}{T}\left[\left(\sum_{i=1}^{T} a_i^2\right) - 2T\mu^2 + T\mu^2\right] \\
&= \frac{1}{T}\sum_{i=1}^{T} a_i^2 - \mu^2.
\end{aligned}$$

The Hypergeometirc Distribution

Example 4. Consider a group of T persons, and let $a_1, \ldots, a_T$ be the heights of these T persons. Suppose that n persons are selected from this group at random without replacement, and let X denote the sum of heights of these n persons. Determine the mean and variance of X.

Solution (approach 1). Then

$$
\begin{aligned}
\mathrm{Cov}[X_i, X_j] &= \frac{2}{T(T-1)} \sum_{i<j} a_i a_j - \frac{1}{T^2} \left(\sum_{i=1}^{T} a_i \right)^2 \\
&= \frac{1}{T^2(T-1)} \left[2T \sum_{i<j} a_i a_j - (T-1) \left(\sum_{i=1}^{T} a_i^2 + 2 \sum_{i<j} a_i a_j \right) \right] \\
&= \frac{1}{T^2(T-1)} \left[\left(\sum_{i=1}^{T} a_i \right)^2 - \sum_{i=1}^{T} a_i^2 - (T-1) \sum_{i=1}^{T} a_i^2 \right] \\
&= \frac{1}{T^2(T-1)} \left[T^2\mu^2 - T^2\sigma^2 - T^2\mu^2 \right] = -\frac{\sigma^2}{T-1}.
\end{aligned}
$$

The Hypergeometirc Distribution

Example 4. Consider a group of T persons, and let $a_1, \ldots, a_T$ be the heights of these T persons. Suppose that n persons are selected from this group at random without replacement, and let X denote the sum of heights of these n persons. Determine the mean and variance of X.

Solution (approach 2). Because $\mathrm{Cov}[X_i, X_j]$ does not depend on i, j as long as $i \neq j$, we have

$$\mathrm{Var}[X] = n\sigma^2 + n(n-1)\mathrm{Cov}[X_1, X_2].$$

Knowing that $\mathrm{Var}[X] = 0$ for $n = T$, we have

$$\mathrm{Cov}[X_1, X_2] = -\frac{1}{T-1}\sigma^2 \quad \Rightarrow \quad \begin{aligned} \mathrm{Var}[X] &= n\sigma^2 - \frac{n(n-1)}{T-1}\sigma^2 \\ &= n\sigma^2\left(\frac{T-n}{T-1}\right). \end{aligned}$$

Exercises

Exercise 1. Suppose Y is the rate (calls per hour) at which calls arrive at a switchboard. Let X be the number of calls during a two-hour period. Suppose the joint probability density function is given by

$$f_{XY}(x, y) = \begin{cases} \dfrac{(2y)^x}{x!} e^{-3y} & \text{for } y > 0 \text{ and } x = 0, 1, \ldots, \\ 0 & \text{otherwise.} \end{cases}$$

(i). Verify that f is a proper joint probability density function.

(ii). Find $P[X = 0]$.

Exercises

Exercise 2. Suppose that X_1 and X_2 are independent random variables, so that

$$X_1 \sim \mathrm{B}(n_1, p), \qquad X_2 \sim \mathrm{B}(n_2, p).$$

For each fixed value of $k (k = 1, 2, \dots, n_1 + n_2)$, prove that the conditional distribution of X_1 given that $X_1 + X_2 = k$ is hypergeometric with parameters $n_1 + n_2, k, n_1$.

Thanks for your attention!