VE401 RC Week4

Wang Yangyang

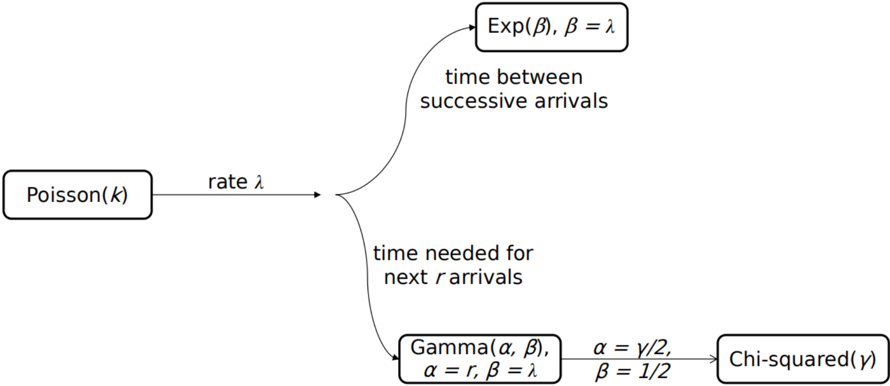
UM-SJTU JI

2022 Spring

Outline

Outline

Connections



Outline

1. **Distributions of Continuous Random Variables**
   - Exponential, Gamma, Chi-Squared Distribution (Done)
   - Normal Distribution
   - Transformation of R.V. and Standardizing
   - Chebyshev's Inequality and Weak Law of Large Number
   - Central Limit Theorem and Normal Approximation
2. **Multivariate Random Variables**
   - Discrete, Continuous Multivariate R.V.
   - Covariance and Correlation
   - Hypergeometric Distribution
   - Transformation of R.V. and Application
3. **Supplementary Materials**
   - Discussion and Exercise

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

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

Normal Distribution

Properties

- Mean.

$$\mathrm{E}[X] = \mu$$

- Variance.

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

- M.G.F.

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

Let's verify the moment-generating function and see what's the takeaway from it.

Normal Distribution

Proof

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

Normal Distribution

Proof

To verify that

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

we use

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

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

Useful Formula

Normal.

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

Gamma.

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

Outline

1. **Distributions of Continuous Random Variables**
   - Exponential, Gamma, Chi-Squared Distribution (Done)
   - Normal Distribution
   - Transformation of R.V. and Standardizing
   - Chebyshev's Inequality and Weak Law of Large Number
   - Central Limit Theorem and Normal Approximation
2. **Multivariate Random Variables**
   - Discrete, Continuous Multivariate R.V.
   - Covariance and Correlation
   - Hypergeometric Distribution
   - Transformation of R.V. and Application
3. **Supplementary Materials**
   - Discussion and Exercise

Transformation of Random Variables

Theorem

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

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

for $y \in \operatorname{ran} \varphi$
and

$$f_Y(y) = 0,$$

for $y \notin \operatorname{ran} \varphi$.

What if not *monotonic and differentiable*? Consider *CDF*.

Transformation of Random Variables

Example

A model for populations of microscopic organisms is exponential growth. Initially, v organisms are introduced into a large tank of water, and let X be the rate of growth. After time t, the population becomes $Y = ve^{Xt}$. Suppose X is unknown and has a continuous distribution

$$f_X(x) = \begin{cases} 3(1-x)^2 & \text{for } 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

What is the distribution of Y?

Transformation of Random Variables

$$f_X(x) = \begin{cases} 3(1-x)^2 & \text{for } 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

$Y = ve^{Xt}$. What is the distribution of Y?

Solution

- Identify and calculate φ^{-1} and $\frac{\mathrm{d}\varphi^{-1}(y)}{\mathrm{d}y}$.
- Substitute x with $\varphi^{-1}(y)$ in the density function of X.

We have $\varphi(x) = ve^{xt}$, and thus

$$\varphi^{-1}(y) = \frac{1}{t}\log\left(\frac{y}{v}\right), \quad \frac{\mathrm{d}\varphi^{-1}(y)}{\mathrm{d}y} = \frac{1}{ty}.$$

$$f_Y(y) = \begin{cases} 3\left(1 - \frac{1}{t}\log\left(\frac{y}{v}\right)\right)^2 \cdot \frac{1}{ty}, & v < y < ve^t \\ 0 & \text{otherwise} \end{cases}$$

Transformation of Random Variables

What if not ***monotonic and differentiable***? Consider ***CDF***.

Example

Consider the continuous random variable X with density

$$f_X(x) = \frac{2/\pi}{e^{-x} + e^x}$$

for $x \in \mathbb{R}$
Find the density of the random variable X^2.

Things become easier using the observation that f_X is an even function.

Transformation of Random Variables

Solution

Note that the function $\varphi : \mathbb{R} \to \mathbb{R}_+ \cup \{0\}, \varphi(x) = x^2$ is not bijective, so we can't simply apply the theorem for transforming random variables from the lecture. Let $y > 0$. Then, using the fact that f_X is even,

$$F_Y(y) = P[Y \leqslant y] = P\left[X^2 \leqslant y\right] = P[-\sqrt{y} \leqslant X \leqslant \sqrt{y}] = \int_{-\sqrt{y}}^{\sqrt{y}} f_X(x)dx$$

$$F_Y(y) = P[Y \leqslant y] = P\left[X^2 \leqslant y\right] = 2\int_0^{\sqrt{y}} f_X(x)dx$$

$$f_Y(y) = F_Y'(y) = 2f_X(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} = \frac{2}{\pi\sqrt{y}} \frac{1}{e^{-\sqrt{y}} + e^{\sqrt{y}}} \quad y > 0$$

For $y \leqslant 0$ we have $F_Y(y) = P[Y \leqslant y] = P\left[X^2 \leqslant y\right] = 0$, so

$$f_Y(y) = 0 \quad y \leqslant 0.$$

Standardizing Normal Distribution

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

CDF

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.

$$\Phi(z) := \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-t^2/2}dt$$

Calculate $P[X < a]$ by

$$\begin{aligned} P[X < a] &= P\left[\frac{X-\mu}{\sigma} < \frac{a-\mu}{\sigma}\right] \\ &= P\left[Z < \frac{a-\mu}{\sigma}\right] \\ &= \Phi\left(\frac{a-\mu}{\sigma}\right) \end{aligned}$$

Outline

1. **Distributions of Continuous Random Variables**
   - Exponential, Gamma, Chi-Squared Distribution (Done)
   - Normal Distribution
   - Transformation of R.V. and Standardizing
   - Chebyshev's Inequality and Weak Law of Large Number
   - Central Limit Theorem and Normal Approximation
2. **Multivariate Random Variables**
   - Discrete, Continuous Multivariate R.V.
   - Covariance and Correlation
   - Hypergeometric Distribution
   - Transformation of R.V. and Application
3. **Supplementary Materials**
   - Discussion and Exercise

Chebyshev's Inequality and Variability

Theorem

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

$$P[|X| \geqslant c] \leqslant \frac{\mathrm{E}\left[|X|^k\right]}{c^k}$$

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

$$P[|X - \mu| \geqslant m\sigma] \leqslant \frac{1}{m^2}$$

or equivalently,

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

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

Law of Large Number

Heuristic Law of Large Number

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

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

Weak Law of Large Number

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 $\varepsilon > 0$,

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

Outline

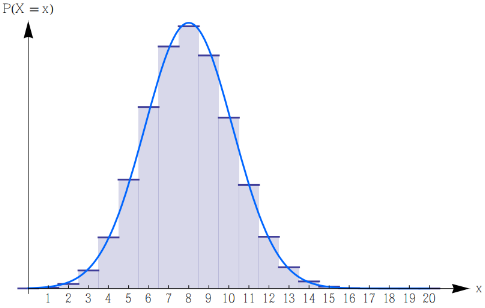
1. **Distributions of Continuous Random Variables**
   - Exponential, Gamma, Chi-Squared Distribution (Done)
   - Normal Distribution
   - Transformation of R.V. and Standardizing
   - Chebyshev's Inequality and Weak Law of Large Number
   - Central Limit Theorem and Normal Approximation
2. **Multivariate Random Variables**
   - Discrete, Continuous Multivariate R.V.
   - Covariance and Correlation
   - Hypergeometric Distribution
   - Transformation of R.V. and Application
3. **Supplementary Materials**
   - Discussion and Exercise

Theorem of De Moivre-Laplace

Theorem

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

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



Normal Approximation of Binomial Distribution

For $y = 0, \ldots, n$

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

where we require that

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

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

Central Limit Theorem

Theorem

Central Limit Theorem. Let (X_i) be a sequence of independent, but not necessarily identical random variables whose third moments exist and satisfy a certain technical condition. Let

$$Y_n = X_1 + \cdots + X_n$$

Then for any $z \in \mathbb{R}$

$$P\left[\frac{Y_n - \mathrm{E}[Y_n]}{\sqrt{\mathrm{Var}[Y_n]}} \leqslant z\right] \xrightarrow{n\to\infty} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2} dx$$

Interpretation

Lyapunov's Central Limit theorem is at the core of the belief by experimentalists that "random error" may be described by the normal distribution.

Outline

Discrete Multivariate Random Variable

Definition

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

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

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

1. $f_{\mathrm{X}}(x) \geqslant 0$ for all $x = (x_1, \ldots, x_n) \in \Omega$ and
2. $\sum_{x \in \Omega} f_{\mathrm{X}}(x) = 1$,

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

Discrete Multivariate Random Variable

Definition

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

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

- *Independent* multivariate random variables:

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

- *Conditional density* of X_1 conditioned on X_2 :

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

Continuous Multivariate Random Variable

Definition

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

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

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

1. $f_{\mathrm{X}}(x) \geqslant 0$ for all $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and
2. $\int_{\mathbb{R}^n} f_{\mathrm{X}}(x) = 1$,

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

Continuous Multivariate Random Variable

Definition

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

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

- *Independent* multivariate random variables:

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

- *Conditional density* of X_1 conditioned on X_2 :

$$f_{X_1|X_2}(x_1) := \frac{f_{X_1X_2}(x_1, x_2)}{f_{X_2}(x_2)} \quad \text{whenever } f_{X_2}(x_2) > 0.$$

Continuous Multivariate Random Variable

Example

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

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

- Verify that f is a proper joint probability density function.
- Find $P[X = 0]$.

Continuous Multivariate Random Variable

Solution

- To verify that f is a proper joint probability density function, we have

$$\begin{aligned}\int_0^\infty \left(\sum_{x=0}^\infty f_{XY}(x,y)\right) dy &= \int_0^\infty \left(\sum_{x=0}^\infty \frac{(2y)^x}{x!}\right) e^{-3y}\, dy \\ &= \int_0^\infty e^{-y}\, dy \\ &= -\left. e^{-y}\right|_0^\infty = 1\end{aligned}$$

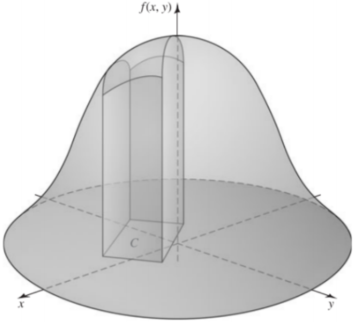
- Plugging in $x = 0$ and integrating with respect to y,

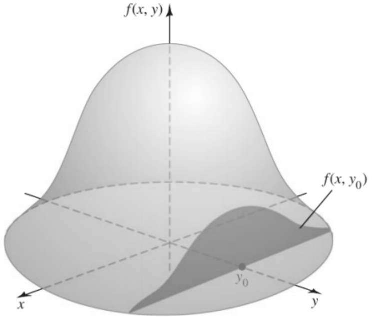
$$P[X=0] = \int_0^\infty f_{XY}(0,y)\mathrm{d}y = \int_0^\infty e^{-3y}\ \mathrm{d}y = \frac{1}{3}.$$

Continuous Multivariate Random Variable

Visualization

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





Continuous Multivariate Random Variable

Definition

For continuous random variables $X_1, \ldots, X_n$, the ***joint cumulative distribution function*** (CDF) is then given by

$$P\left[X_1 \leqslant a_1, \ldots, X_n \leqslant a_n\right] = \int_{-\infty}^{a_1} \cdots \int_{-\infty}^{a_n} f_X(x) \mathrm{d}x_1 \ldots \mathrm{d}x_n$$

Example

Suppose X and Y are random variables that take values in the intervals $0 \leqslant X \leqslant 2$ and $0 \leqslant Y \leqslant 2$. 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 Variable

Solution

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

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

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

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

Outline

Expectation

- Discrete.

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

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

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

- Continuous.

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

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

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

Covariance

Definition

Definition. For a multivariate random variable X, the ***covariance matrix*** Var[X] is given by

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

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

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

and

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

Covariance

Properties

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

- X and Y are independent $\Rightarrow \mathrm{Cov}[X, Y] = 0$.
  The Converse is not True!
- $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\,\mathrm{Cov}[X, Y]$, and more generally,

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

  if $\mathrm{Var}\left[X_i\right] < \infty$ for $i = 1, \ldots, n$
- $\mathrm{Cov}[X, Y+Z] = \mathrm{Cov}[X, Y] + \mathrm{Cov}[X, Z]$
  $\mathrm{Cov}[X, Y-Z] = \mathrm{Cov}[X, Y] - \mathrm{Cov}[X, Z]$
- $\mathrm{Cov}[X, X] = \mathrm{Var}[X]$

Correlation

Definition

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

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

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

- $-1 \leqslant \rho_{XY} \leqslant 1$,
- $|\rho_{XY}| = 1$ iff there exist $\beta_0, \beta_1 \in \mathbb{R}$ such that

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

Correlation

Note

- Uncorrelated does not mean independent,
- Correlation coefficient only measures linear relationships.

Example

Two random variables X and Y. X follows a uniform distribution $U(-1,1)$ and $Y = X^2$. Find $\mathrm{Cov}(X,Y)$.

$$
\begin{aligned}
\mathrm{Cov}(X,Y) &= \mathrm{Cov}\left(X, X^2\right) \\
&= \mathrm{E}\left((X-\mathrm{E}(X))\left(X^2-\mathrm{E}\left(X^2\right)\right)\right) \\
&= \mathrm{E}\left(X^3 - X^2\mathrm{E}(X) - X\mathrm{E}\left(X^2\right) + \mathrm{E}(X)\mathrm{E}\left(X^2\right)\right) \\
&= \mathrm{E}\left(X^3\right) - \mathrm{E}\left(X^2\right)\mathrm{E}(X) - \mathrm{E}(X)\mathrm{E}\left(X^2\right) + \mathrm{E}(X)\mathrm{E}\left(X^2\right) \\
&= \int_{-1}^{1} \frac{1}{2}x^3 \, \mathrm{d}x - \int_{-1}^{1} \frac{1}{2}x^2 \, \mathrm{d}x \cdot \int_{-1}^{1} \frac{1}{2}x \, \mathrm{d}x = 0
\end{aligned}
$$

Correlation

Example

There is one more example. Suppose X has a standard normal distribution. Let W follows a distribution where $W = 1$ or $W = -1$, each with probability $1/2$, and assume W is independent of X. Let $Y = WX$. Then

- X and Y are uncorrelated;
- both have the same normal distribution; and
- X and Y are not independent.

To see that X and Y are uncorrelated, by the independence of W from X, one has

$$\operatorname{cov}(X, Y) = \mathrm{E}(XY) - 0 = \mathrm{E}\left(X^2 W\right) = \mathrm{E}\left(X^2\right)\mathrm{E}(W) = \mathrm{E}\left(X^2\right) \cdot 0 = 0$$

To see that X and Y are not independent, observe that $|Y| = |X|$

The Fisher Transformation

Definition

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

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

We say that X and Y are

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

The Bivariate Normal Distribution

The density function of ***Bivariate Normal Distribution***:

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

- $-1 < \rho < 1$
- $\mu_X = \mathrm{E}[X]$, $\sigma_X^2 = \operatorname{Var} X$ (and similarly for Y).
- $\rho = \rho_{XY}$ is indeed the correlation coefficient of X and Y.
- X and Y are independent $\iff \rho = 0$

Outline

Hypergeometric Distribution

Definition

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

$$f_X(x) = \frac{\begin{pmatrix} r \\ x \end{pmatrix}\begin{pmatrix} N-r \\ n-x \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}$$

Interpretation

- $f_X(x)$ is the probability of getting x red 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 $\frac{r}{N}$.

Hypergeometric Distribution

Property

- Expectation

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

- Variance

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

The ***binomial distribution*** may be used to approximate the hypergeometric distribution if n/N is small (less than 0.05).

Hypergeometric Mean

Transform to Bernoulli trials $(X_1, \ldots, X_n)$.
The Bernoulli trials are identical with $p_k = \frac{r}{N}$, i.e.,

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

and so on.

Hypergeometric Variance

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

We need to calculate $\mathrm{Cov}[X_i, X_j] = \mathrm{E}[X_i X_j] - \mathrm{E}[X_i]\,\mathrm{E}[X_j]$. For this, we note that $X_i X_j$ is also a Bernoulli variable, since

$$X_i X_j = \begin{cases} 1 & \text{if } X_i = 1 \text{ and } X_j = 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\mathrm{E}[X_i X_j] = p_{ij} := P[X_i = 1 \text{ and } X_j = 1] = \frac{r}{N} \cdot \frac{r-1}{N-1}$$

$$\mathrm{Var}[X_i] = \frac{r}{N}\left(1 - \frac{r}{N}\right), \quad \mathrm{Cov}[X_i, X_j] = -\frac{1}{N} \cdot \frac{r(N-r)}{N(N-1)}$$

Since there are $\binom{n}{2}$ pairs (i, j) with $i < j$, finally gives

$$\mathrm{Var}[X] = n\frac{r}{N}\frac{N-r}{N}\frac{N-n}{N-1}$$

Hypergeometric Distribution

Theorem

Suppose Y has a binomial distribution with parameters $n \in \mathbb{N} \backslash \{0\}$ 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 $x = 0, \ldots, n$,

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

Hypergeometric Sample

Exercise

Consider a group of T people, and let $a_1, \ldots, a_T$ with mean μ and variance σ^2 denote the heights of these T people. Suppose that n people 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,

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

Hypergeometric Sample

Solution

How to solve Covariance?

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

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

Outline

Transformation of Random Variables

Theorem

Let $(\boldsymbol{X}, f_{\boldsymbol{X}})$ be a continuous multivariate random variable and let $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ be a differentiable, bijective map with inverse φ^{-1}. Then $\boldsymbol{Y} = \varphi \circ \boldsymbol{X}$ is a continuous multivariate random variable with density

$$f_Y(y) = f_X \circ \varphi^{-1}(y) \cdot \left|\det D\varphi^{-1}(y)\right|,$$

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

- $f_Y(y) = 0$ for $y \notin \operatorname{ran} \varphi$.
- When the map is not ***strictly monotonic***, we usually consider ***Cumulative Density Function***.
- M.G.F. may help sometimes.

Quotient of Normal: Cauchy

Lemma

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

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

Theorem

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

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

Quotient of Normal: Cauchy

Proof

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

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

Then the joint density function is given by

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

Then the marginal of U is calculated as

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

Root Sum of Normal Square: Chi

Definition

χ_n is a ***chi random variable*** with n ***degrees of freedom***,

$$\chi_n = \sqrt{\sum_{i=1}^{n} Z_i^2}$$

where $Z_1, \ldots, Z_n$ are ***independent standard normal*** random variables.

$$f_{\chi_n}(y) = \frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} y^{n-1} e^{-y^2/2} \quad (y > 0)$$

Interpretation

A chi random variable represents the sum of the root squares (distance) of independent standard normal variables.

Sum of Normal Square: Chi-Squared

Definition

χ_n^2 is a *chi-squared random variable* with n *degrees of freedom*,

$$\chi_n^2 = \sum_{i=1}^{n} Z_i^2$$

where $Z_1, \ldots, Z_n$ are *independent standard normal* random variables.

$$f_{\chi_n^2}(y) = \frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} y^{n/2-1} e^{-y/2} \quad (y > 0)$$

Interpretation

A chi-squared random variable represents the sum of the squares of independent standard normal variables.

Sum of Normal: Normal

Theorem

If the random variables $X_1, \ldots, X_k$ are independent and if X_i follows *normal distribution* with mean μ_i and variance σ_i^2, where $i = 1, \ldots, k$, then $X = X_1 + \cdots + X_k$ follows normal distribution with

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

Proof

Using M.G.F., we have

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

Sum of Chi-Squared: Chi-Squared

Lemma

Let $\chi^2_{\gamma_1}, \ldots, \chi^2_{\gamma_n}$ be n independent random variables following chi-squared distributions with $\gamma_1, \ldots, \gamma_n$ degrees of freedom, respectively. Then

$$\chi^2_\alpha := \sum_{k=1}^{n} \chi^2_{\gamma_k}$$

is a ***chi-squared random variable*** with $\alpha = \sum_{k=1}^{n} \gamma_k$ degrees of freedom.

Outline

1. Trigonometric Transformation

Exercise

Let (X, Y) be a continuous bivariate random variable with density $f_{XY} : S \to \mathbb{R}^2$ given by

$$f_{XY}(x, y) = \begin{cases} c \cdot \left(x^2 + y^2\right) & \text{for } x^2 + y^2 \leqslant 1 \\ 0 & \text{otherwise} \end{cases}$$

where $c \in \mathbb{R}$ is a suitable constant.

1. Determine the constant c.
2. Find $\mathrm{E}[X]$ and $\mathrm{E}[Y]$.
3. Find $\mathrm{Var}[X]$ and $\mathrm{Var}[Y]$.
4. Find the correlation coefficient ρ_{XY}.

2. Not Bijective Map Transformation

Exercise

(*Univariate* R.V.) Let X be a continuous *uniformly distributed* random variable on $[-1, 1]$. Does X^2 also follow uniform distribution?

Exercise

(*Multivariate* R.V.) Let (X, Y) be a continuous bivariate random variable with density $f_{XY} : S \to \mathbb{R}^2$ given by

$$f_{XY}(x, y) = \begin{cases} \frac{2}{\pi} \cdot \left(x^2 + y^2\right) & \text{for } x^2 + y^2 \leqslant 1 \\ 0 & \text{otherwise} \end{cases}$$

Find the density of $U = X/Y$.

3. Proof for Independence

Exercise

Let X, Y be independent random variables such that $X, Y \sim N(\mu, \sigma^2)$, show that $X+Y$ and $X-Y$ are independent.

Can we simply state that

$$\mathrm{Cov}(X+Y, X-Y) = \mathrm{Cov}(X,X) + \mathrm{Cov}(X,Y) - \mathrm{Cov}(X,Y) - \mathrm{Cov}(Y,Y)$$

$$= \mathrm{Var}(X) - \mathrm{Var}(Y) = \sigma^2 - \sigma^2 = 0$$

4. Chi and Chi-Squared Distribution

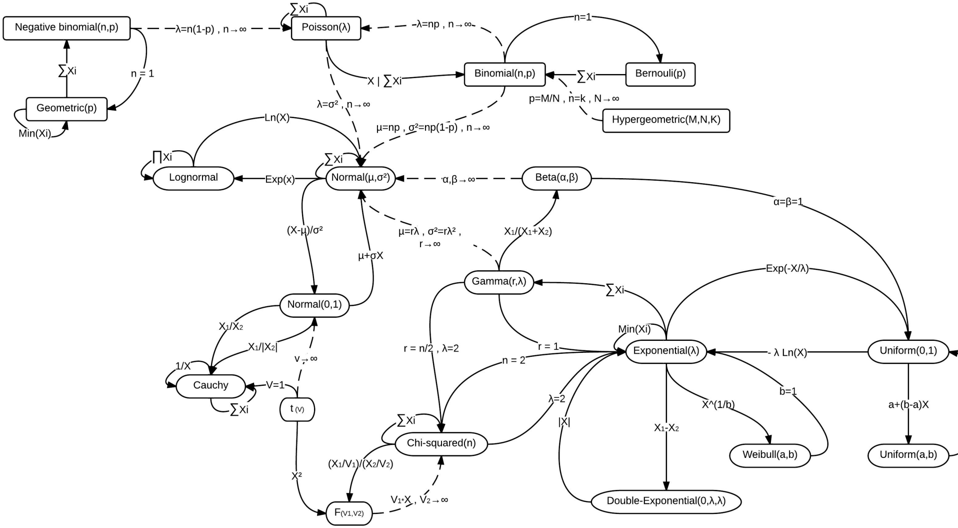
Discussion

How to derive Chi and *Chi-Squared* distribution from *Normal* Distribution?

Exercise

We can also derive *Chi-Squared* following *Uniform*-*Exponential*-*Gamma* thread.

1. Gamma distribution with parameters $\alpha = r, \beta = \lambda$ has M.G.F. $m_X(t) = \frac{1}{(1-t/\beta)^\alpha}$. What is the M.G.F. for Chi-squared distribution with γ degrees of freedom?
2. If the random variables $X_1, X_2, \dots, X_n$ are independent and follow the uniform distribution $U(0,1)$. Find the distribution of the random variable $Z = \sum_{i=1}^{n} Y_i$, where $Y = -2 \ln X$.



Negative binomial(n,p)
λ=n(1-p) , n→∞
Poisson(λ)
∑Xi
λ=np , n→∞
n=1
Binomial(n,p)
Bernouli(p)
∑Xi
n = 1
Geometric(p)
Min(Xi)
X | ∑Xi
∑Xi
p=M/N , n=k , N→∞
Hypergeometric(M,N,K)
λ=σ² , n→∞
Ln(X)
μ=np , σ²=np(1-p) , n→∞
∏Xi
Lognormal
Exp(x)
∑Xi
Normal(μ,σ²)
α,β→∞
Beta(α,β)
α=β=1
(X-μ)/σ²
μ=rλ , σ²=rλ² , r→∞
$X_1/(X_1+X_2)$
μ+σX
Gamma(r,λ)
∑Xi
Exp(-X/λ)
Normal(0,1)
X_1/X_2
$X_1/|X_2|$
Min(Xi)
r = n/2 , λ=2
r = 1
Exponential(λ)
- λ Ln(X)
Uniform(0,1)
1/X
v→∞
n = 2
Cauchy
V=1
∑Xi
t (V)
λ=2
b=1
|X|
X^(1/b)
a+(b-a)X
∑Xi
Chi-squared(n)
X_1-X_2
Weibull(a,b)
Uniform(a,b)
$(X_1/V_1)/(X_2/V_2)$
X²
V_1*X , V_2→∞
F(V1,V2)
Double-Exponential(0,λ,λ)

End

Credit to Zhanpeng Zhou (TA of SP21)
Credit to Fan Zhang (TA of SU21)
Credit to Jiawen Fan (TA of SP21)
Credit to Zhenghao Gu (TA of SP20)
https://www.johndcook.com/blog/distribution_chart/