VE401, Probabilistic Methods in Eng.

Recitation Class - Mid Exam Review

Zhanpeng Zhou

UMJI-SJTU Joint Institute

March 26, 2021

Mail: zzp1012@sjtu.edu.cn / WeChat: zzp01012

Table of contents

Continuous Random Variables

Definition. Let S be a sample space. A ***continuous random variable*** is a map $X : S \to \mathbb{R}$ together with a function $f_X : \mathbb{R} \to \mathbb{R}$ with the properties that

(i) $f_X(x) \geq 0$ for all $x \in \mathbb{R}$ and

(ii) $\displaystyle\int_{-\infty}^{\infty} f_X(x)\mathrm{d}x = 1.$

The integral of f_X is interpreted as the probability that X assumes values x in a given range, i.e.,

$$P[a \leq X \leq b] = \int_a^b f_X(x)\mathrm{d}x.$$

The function f_X is called the ***probability density function*** of random variable X.

Cumulative Distribution

Definition. Let (X, f_X) be a continuous random variable. The cumulative distribution function for X is defined by $F_X : \mathbb{R} \to \mathbb{R}$,

$$F_X(x) := P[X \leq x] = \int_{-\infty}^{x} f_X(y)\mathrm{d}y.$$

By the fundamental theorem of calculus, we can obtain the density function from F_X by

$$f_X(x) = F_X'(x).$$

Expectation, Variance, and M.G.F.

- Expectation.

$$\mathrm{E}[X] := \int_{\mathbb{R}} x \cdot f_X(x)\mathrm{d}x.$$

- Variance.

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

- Moment-generating function.

$$m_X(t) = \mathrm{E}[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f_X(x)\mathrm{d}x.$$

Note. All previous properties about expectation, variance and M.G.F. hold for continuous random variables.

Location of Continuous Distributions

Definitions.

- The ***median*** M_X is defined by $P[X \le M_X] = 0.5$.
- The ***mean*** is given by $\mathrm{E}[X]$.
- The ***mode*** x_0, is the location of the maximum of f_X.

Exponential Distribution

Definition. A continuous random variable (X, f_β) a ***exponential distribution*** with parameter β if the probability density function is defined by

$$f_\beta(X) = \begin{cases} \beta e^{-\beta x}, & x > 0, \\ 0, & x \le 0. \end{cases}$$

Interpretation. The time between successive arrivals of a Poisson process with rate λ follows exponential distribution with parameter $\beta = \lambda$. (Recall $P[T > t] = e^{-\beta t}$.)

Note. Memoryless property:

$$P[X > x + s | X > x] = P[X > s].$$

Exponential Distribution

Mean, variance and M.G.F.

- Mean.

$$\mathrm{E}[X] = \frac{1}{\beta}.$$

- Variance.

$$\mathrm{Var}[X] = \frac{1}{\beta^2}.$$

- M.G.F.

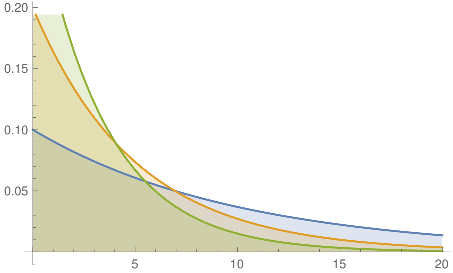
$$m_X : (-\infty, \beta) \to \mathbb{R}, \qquad m_X(t) = \frac{1}{1 - t/\beta}.$$

Exponential Distribution

Plots.

In[∘]:=
```
Plot[Table[PDF[ExponentialDistribution[β], x], {β, {0.1, 0.2, 0.3}}] // Evaluate, {x, 0, 20},
 Filling → Axis]
```

Out[∘]=



Exponential Distribution

Example 1. n light bulbs are burning simultaneously and independently, and the lifetime for each bulb follows the $\text{Exp}(\beta)$.

(i). What is the distribution of the length of time Y_1 until the first failure in one of the n bulbs?

(ii). What is the distribution of the length of time Y_2 after the first failure until a second bulb fails?

Exponential Distribution

Example 1. n light bulbs are burning simultaneously and independently, and the lifetime for each bulb follows the $\text{Exp}(\beta)$.

(i) What is the distribution of the length of time Y_1 until the first failure in one of the n bulbs?

(ii) What is the distribution of the length of time Y_2 after the first failure until a second bulb fails?

Solution (i). Suppose random variables $X_1, \ldots, X_n$ satisfies that $X_i \sim \text{Exp}(\beta)$, and $Y_1 = \min\{X_1, \ldots, X_n\}$. Then for any $t > 0$,

$$\begin{aligned} P[Y_1 > t] &= P[X_1 > t, \ldots, X_n > t] \\ &= P[X_1 > t] \times \cdots \times P[X_n > t] \\ &= e^{-n\beta t}, \end{aligned}$$

indicating an exponential distribution with parameter $n\beta$. What about (ii)?

Gamma Distribution

Definition. Let $\alpha, \beta \in \mathbb{R}, \alpha, \beta > 0$. A continuous random variable $(X, f_{\alpha,\beta})$ follows a ***gamma distribution*** with parameters α and β if the probability density function is given by

$$f_{\alpha,\beta}(x) = \begin{cases} \dfrac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, & x > 0, \\ 0, & x \le 0. \end{cases}$$

where $\Gamma(\alpha) = \int_0^\infty z^{\alpha-1} e^{-z} \mathrm{d}z, \alpha > 0$ is the Euler gamma function.

Interpretation. The time needed for the next r arrivals in a Poisson process with rate λ follows a Gamma distribution with parameters $\alpha = r, \beta = \lambda$.

Gamma Distribution

Mean, variance and M.G.F.

- Mean.

$$\mathrm{E}[X] = \frac{\alpha}{\beta}.$$

- Variance.

$$\mathrm{Var}[X] = \frac{\alpha}{\beta^2}.$$

- M.G.F.

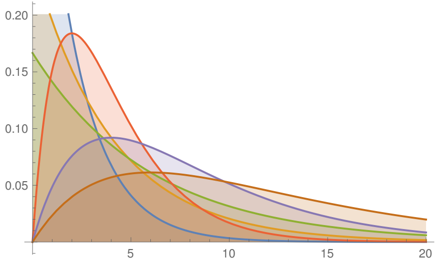
$$m_X : (-\infty, \beta) \to \mathbb{R}, \qquad m_X(t) = \frac{1}{(1 - t/\beta)^\alpha}.$$

Gamma Distribution

Plots.

```
In[∘ ]:= Plot[Table[PDF[GammaDistribution[α, β], x], {α, {1, 2}}, {β, {2, 4, 6}}] // Evaluate, {x, 0, 20},
  Filling → Axis]
```

Out[∘]=



Chi-Squared Distribution

Definition. Let $\gamma \in \mathbb{N}$. A continuous random variable (X_γ^2, f_X) follows a ***chi-squared distribution*** with γ degrees of freedom if the probability density function is given by

$$f_\gamma(x) = \begin{cases} \dfrac{1}{2^{\gamma/2}\Gamma(\gamma/2)} x^{\gamma/2-1} e^{-x/2}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

which is a gamma distribution with $\alpha = \gamma/2, \beta = 1/2$. Therefore,

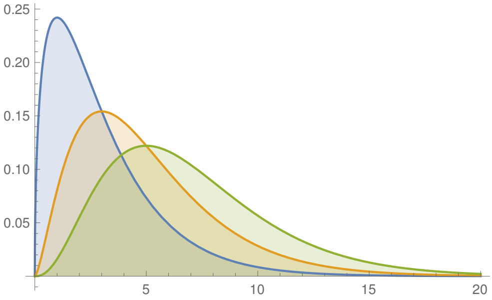
$$\mathrm{E}[X_\gamma^2] = \gamma, \qquad \mathrm{Var}[X_\gamma^2] = 2\gamma.$$

Chi-Squared Distribution

Plots.

```
In[ ]:= Plot[Table[PDF[ChiSquareDistribution[γ], x], {γ, {3, 5, 7}}] // Evaluate, {x, 0, 20},
         Filling → Axis]
```

Out[ ]=



Distributions based on of Poisson Process

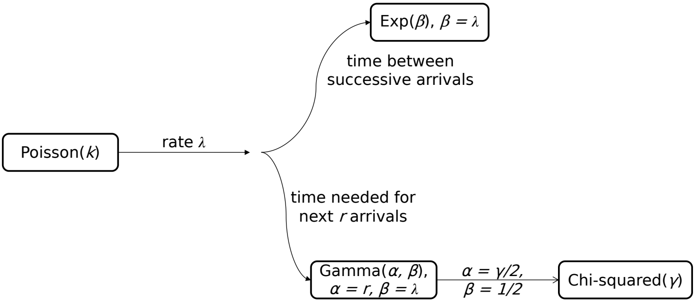


Figure: Connections of distributions based on Poisson process.

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.

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

- 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.

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

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.
  1. Normal.

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

  2. Gamma.

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

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***.

Cumulative Distribution Function

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

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

The Chebyshev's Inequality

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

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

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

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

Theorem of De Moivre-Laplace

Theorem of De Moivre-Laplace. 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

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

Normal Approximation of Binomial Distribution

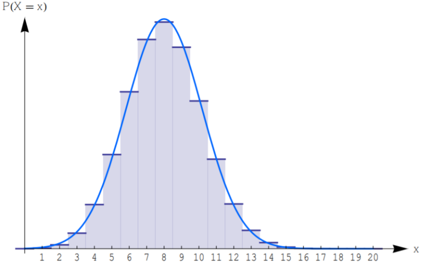
For $y = 0, \ldots, n$,

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

where we require that

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

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. 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 $z \in \mathbb{R}$

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

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 $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ 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 $X_k, k = 1, \ldots, n$:

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

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

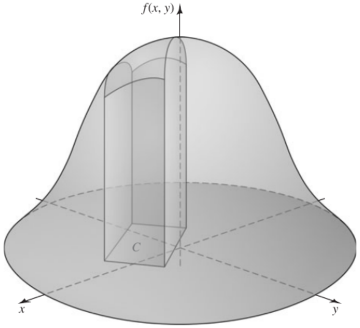
$$f_{\mathbf{X}}(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n).$$

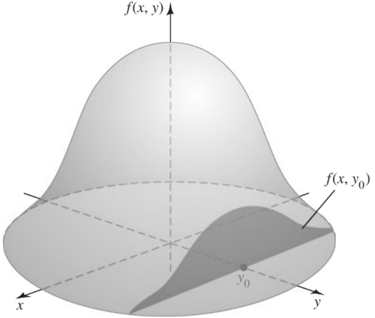
- ***Conditional density*** of X_1 conditioned on X_2:

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

Continuous Multivariate Random Variables

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 $X_1, \dots, X_n$, the joint cumula-tive distribution function is then given by

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

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?

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

- Discrete.

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

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

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

- Continuous.

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

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

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

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

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

and

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

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

$$\mathrm{E}[XY] = \mathrm{E}[X^3] = \mathrm{E}[X] = 0,$$

and thus

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

Pearson Correlation Coefficient

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

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

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

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

We say that X and Y are

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

Bivariate Normal Distribution

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

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

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

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

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

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

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

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

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

and

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

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}.

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,

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

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 $k = k_1 + \cdots + k_n$.

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

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

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

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

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

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

Exercise 1.

Exercise 1. Suppose Keven plays a game where he has probability p to win in each play. When he wins, his fortune is doubled, and when he loses, his fortune is cut in half. If he begins playing with a given fortune $c > 0$, what is the expected value of his fortune after n independent plays?

Exercise 1. Sol

Solution. Define a random variable $X_i, i = 1, \ldots, n$, where $X_i = 2$ if Keven's fortune is doubled on the i-th play and $X_i = 1/2$ if his fortune is cut in half on the i-th play. Then

$$\mathrm{E}[X_i] = 2p + \frac{1}{2}(1-p) = \frac{1}{2}(1+3p).$$

Because the plays are independent from each other, the expected final fortune is

$$c\mathrm{E}[X_1 \cdots X_n] = c \times \mathrm{E}[X_1] \cdots \mathrm{E}[X_n] = \frac{c}{2^n}(1+3p)^n.$$

Exercise 2.

Exercise 2. Suppose that a book with n pages contains on the average λ misprints per page. What is the probability that there will be at least m pages which contain more than k misprints?

Exercise 2. Sol I

Solution. Let Y denote the number of misprints on a given page. Then the probability p that a given page will contain more than k misprints is

$$p = P[Y > k] = \sum_{i=k+1}^{\infty} f(i;\lambda) = \sum_{i=k+1}^{\infty} \frac{\lambda^i e^{-\lambda}}{i!},$$

and

$$1 - p = \sum_{i=0}^{k} f(i;\lambda) = \sum_{i=0}^{k} \frac{\lambda^i e^{-\lambda}}{i!}.$$

Let X denote the number of pages among n pages in the book, where there are more than k misprints. Then for $x = 0, 1, \ldots, n$,

$$P[X = x] = \binom{n}{x} p^x (1-p)^{n-x},$$

Exercise 2. Sol II

which gives

$$P[X \geq m] = \sum_{x=m}^{n} \binom{n}{x} p^x (1-p)^{n-x},$$

where p is given by the formula above.

Exercise 3.

Exercise 3. Suppose that a certain system contains three components C_1, C_2, C_3 that function independently of each other and are connected as series, so that the system fails as soon as one of the components fails. Suppose that the length of life of the C_1, C_2, C_3 has the exponential distribution with parameters

$$\beta_1 = 0.001, \qquad \beta_2 = 0.003, \qquad \beta_3 = 0.006,$$

respectively, all measured in hours. Determine the probability that the system will not fail before 100 hours.

Exercise 3. Sol

Solution. Let Y denote the length of life of the system, and X_1, X_2, X_3 are the lifetime of each component. Then for any $y > 0$,

$$\begin{aligned} P[Y > y] &= P[X_1 > y, X_2 > y, X_3 > y] \\ &= P[X_1 > y]P[X_2 > y]P[X_3 > y] \\ &= e^{-(\beta_1+\beta_2+\beta_3)y}. \end{aligned}$$

Therefore,

$$P[Y > 100] = e^{-100\times 0.01} = \frac{1}{e}.$$

Exercise 4.

Exercise 4. 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

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

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

Exercise 4. Sol.

Solution.

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

$$\int_0^\infty \left(\sum_{x=0}^\infty f_{XY}(x,y) \right) \mathrm{d}y = \int_0^\infty \left(\sum_{x=0}^\infty \frac{(2y)^x}{x!} \right) e^{-3y} \mathrm{d}y$$
$$= \int_0^\infty e^{-y} \mathrm{d}y$$
$$= -e^{-y} \Big|_0^\infty = 1.$$

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

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

Exercise 5.

Exercise 5. Considering the table on the lecture slide pp.193:

x/y	0	1	2	3	4	$f_X(x)$
0	1	1	4	1	1	$8/36$
1	1	2	1	2	1	$7/36$
2	2	1	1	1	2	$7/36$
3	2	1	1	1	2	$7/36$
4	1	2	1	2	1	$7/36$
$f_Y(y)$	$7/36$	$7/36$	$8/36$	$7/36$	$7/36$	1

We have already known that X and Y are not independent.
Try to calculate ρ_{XY}.

Exercise 5. Sol I

Solution

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

where

$$\begin{aligned}
\mathrm{E}[XY] &= \sum_{x,y\in\Omega} xy \cdot f_{XY}(x,y) \\
&= 1\cdot\tfrac{2}{36} + 2\cdot\tfrac{1}{36} + 3\cdot\tfrac{2}{36} + 4\cdot\tfrac{1}{36} + 2\cdot\tfrac{1}{36} + 4\cdot\tfrac{1}{36} + 6\cdot\tfrac{1}{36} + 8\cdot\tfrac{2}{36} + \\
&\quad 3\cdot\tfrac{1}{36} + 6\cdot\tfrac{1}{36} + 9\cdot\tfrac{1}{36} + 12\cdot\tfrac{2}{36} + 4\cdot\tfrac{2}{36} + 8\cdot\tfrac{1}{36} + 12\cdot\tfrac{2}{36} + 16\cdot\tfrac{1}{36} \\
&= \frac{35}{9}
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}[X] &= \sum_{x,y\in\Omega} x \cdot f_{XY}(x,y) \\
&= \sum_{x\in\Omega} x \cdot f_X(x) \\
&= 1\cdot\frac{7}{36} + 2\cdot\frac{7}{36} + 3\cdot\frac{7}{36} + 4\cdot\frac{7}{36} = \frac{35}{18}
\end{aligned}$$

Exercise 5. Sol II

$$
\begin{aligned}
\mathrm{E}[Y] &= \sum_{x,y\in\Omega} y \cdot f_{XY}(x,y) \\
&= \sum_{y\in\Omega} y \cdot f_Y(y) \\
&= 1\cdot\frac{7}{36} + 2\cdot\frac{8}{36} + 3\cdot\frac{7}{36} + 4\cdot\frac{7}{36} = 2
\end{aligned}
$$

Hence,

$$
\mathrm{Cov}[X, Y] = \frac{35}{9} - \frac{35}{18}\cdot 2 = 0
$$

Therefore,

$$
\rho_{XY} = \frac{\mathrm{Cov}[X, Y]}{\sqrt{(\mathrm{Var}[X])(\mathrm{Var}[Y])}} = 0
$$