

Probability Exam Notes

Ryan Lin

2024-05-01

1 Discrete Distributions

1.1 Discrete Uniform Distribution

The probability \mathbb{P} of rolling x on an n -sided die where the probability of each result being rolled is equal.

$$\mathbb{P}(X = x) = \frac{1}{n}$$

The probability of each roll being equal ($\mathbb{P}(X = x)$ for all x) and the sum of all the probabilities being equal to 1 ($\sum \mathbb{P}(X = x) = 1$) implies the probability of each roll is $\frac{1}{n}$.

1.2 Bernoulli Distribution

The probability (\mathbb{P}) of a student passing (b) his test where the probability of a student passing his test is p .

$$\mathbb{P}(X = b) = \begin{cases} p, & \text{if } b = 1 \\ 1 - p, & \text{if } b = 0 \end{cases}$$

This is pretty straightforward; there is nothing more to say here.

1.3 Binomial Distribution

The probability (\mathbb{P}) of any m out of n students passing their tests where the probability of 1 student passing his test is p .

$$\mathbb{P}(X = m) = \binom{n}{m} p^m (1 - p)^{n-m}$$

A combinatorial explanation:

- The probability of 1 student passing his test: p
- The probability of m students simultaneously passing their tests: p^m
- The probability of the remaining students out of n failing their tests: $(1 - p)^{n-m}$
- The probability of a certain m out of n passing and the remaining $n - m$ students failing: $p^m (1 - p)^{n-m}$
- The number of combinations of any m out of n students passing (and consequently $n - m$ failing): $\binom{n}{m} (= \binom{n}{n-m})$
- The probability of any m out of n students passing their tests and the remaining $n - m$ students failing: $\binom{n}{m} p^m (1 - p)^{n-m}$

1.4 Geometric Distribution (1)

Student passing his test within k tries. The probability (\mathbb{P}) of winning the jackpot in k plays where the probability of winning the jackpot in 1 play is p .

$$\mathbb{P}(X = k) = (1 - p)^{k-1}p$$

A combinatorial explanation:

- The probability of winning the jackpot in 1 play: p
- The probability of losing the jackpot $k - 1$ times: $(1 - p)^{k-1}$
- The probability of losing the jackpot $k - 1$ times and winning the jackpot once in k plays: $(1 - p)^{k-1}p$

1.5 Geometric Distribution (2)

The probability (\mathbb{P}) of winning the jackpot within k plays where the probability of winning the jackpot in 1 play is p

$$\mathbb{P}(X \leq k) = \sum_{i=1}^k (1 - p)^{i-1}p$$

A combinatorial explanation:

- The probability of winning the jackpot in 1 play: $(1 - p)^{1-1}p$
- The probability of winning the jackpot in 2 plays: $(1 - p)^{2-1}p$
- \vdots
- The probability of winning the jackpot in k plays: $(1 - p)^{k-1}p$
- The probability of winning the jackpot in k plays or less: $(1 - p)^{1-1}p + (1 - p)^{2-1}p + \dots + (1 - p)^{k-1}p = \sum_{i=1}^k (1 - p)^{i-1}p$

1.6 Poisson Distribution

The probability (\mathbb{P}) of m out of infinite atoms decaying where the expected number of atoms decaying is λ .

$$\mathbb{P}(X = m) = \frac{\lambda^m}{m!} e^{-\lambda}$$

An algebraic explanation:

There exist n atoms and the probability of each atom decaying is p . So the expected number of atoms decaying is $\lambda = np$. This equation can be rearranged into $p = \frac{\lambda}{n}$.

The probability of m out of n atoms decaying is the binomial distribution $\binom{n}{m} p^m (1 - p)^{n-m}$. So the probability of m out of infinite atoms decaying is

$$\begin{aligned} \lim_{n \rightarrow \infty} \binom{n}{m} p^m (1 - p)^{n-m} &= \lim_{n \rightarrow \infty} \frac{n!}{m!(n-m)!} \left(\frac{\lambda}{n}\right)^m \left(1 - \frac{\lambda}{n}\right)^{n-m} \\ &= \lim_{n \rightarrow \infty} \frac{1}{m!} \frac{n(n-1) \cdots (n-m+1)}{n^m} \lambda^m \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-m} \\ &= \frac{1}{m!} (1) \lambda^m (e^{-\lambda}) (1)^{-m} \\ &= \frac{\lambda^m}{m!} e^{-\lambda} \end{aligned}$$

2 Continuous Distributions

2.1 Exponential Distribution

The probability (\mathbb{P}) of some atom decaying within t time where the expected number of atoms decaying in t unit time is λ .

$$\begin{aligned}m_Y(t) &= \mathbb{P}(Y < t) = 1 - \lambda e^{-\lambda t} \\ \rho_Y(t) &= \frac{d}{dt} m_Y(t) = \lambda e^{-\lambda t}\end{aligned}$$

An algebraic explanation:

The probability of an atom decaying in t unit times is p and the total number of atoms is n so the expected number of atoms that decay in t unit times is $p \cdot n$. The expected number of atoms that decay in a unit time is λ and the number of unit times is n so the expected number of atoms that decay in t unit times is $\lambda \cdot t$. It follows that $p \cdot n = \lambda \cdot t \Rightarrow p = \frac{\lambda t}{n}$.

Define a random variable X to be the number of atoms out of n decaying within t time where the probability of 1 atom decaying in t time is p .

The probability mass of m out of n atoms decaying within t time:

$$\begin{aligned}\mathbb{P}(X = m) &= \binom{n}{m} p^m (1-p)^{n-m} \\ &= \binom{n}{m} \left(\frac{\lambda t}{n}\right)^m \left(1 - \frac{\lambda t}{n}\right)^{n-m}\end{aligned}$$

The probability of m out of infinite atoms decaying within t time:

$$\begin{aligned}\mathbb{P}(X = m) &= \lim_{n \rightarrow \infty} \left(\binom{n}{m} \left(\frac{\lambda t}{n}\right)^m \left(1 - \frac{\lambda t}{n}\right)^{n-m} \right) \\ &= \lim_{n \rightarrow \infty} \left(\frac{n!}{m!(n-m)!} \left(\frac{\lambda t}{n}\right)^m \left(1 - \frac{\lambda t}{n}\right)^{n-m} \right) \\ &= \lim_{n \rightarrow \infty} \left(\frac{n(n-1) \cdots (n-m+1)}{m!} \frac{(\lambda t)^m}{n^m} \left(1 - \frac{\lambda t}{n}\right)^n \left(1 - \frac{\lambda t}{n}\right)^{-m} \right) \\ &= \lim_{n \rightarrow \infty} \left(\frac{(\lambda t)^m}{m!} \frac{n(n-1) \cdots (n-m+1)}{n^m} \right) \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda t}{n}\right)^n \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda t}{n}\right)^{-m} \\ &= \frac{(\lambda t)^m}{m!} (1)(e^{-\lambda t})(1) \\ &= \frac{(\lambda t)^m}{m!} e^{-\lambda t}\end{aligned}$$

The probability of at least 1 atom decaying within t time:

$$\begin{aligned}\mathbb{P}(X \neq 0) &= 1 - \mathbb{P}(X = 0) \\ &= 1 - \frac{(\lambda t)^0}{0!} e^{-\lambda t} \\ &= 1 - e^{-\lambda t}\end{aligned}$$

Define a random variable Y to be the time t within which some atom decays.

The probability mass function of some atom decaying within t time:

$$\begin{aligned} m_Y(t) &= \mathbb{P}(Y = t) \\ &= \mathbb{P}(X \neq 0) \\ &= 1 - e^{-\lambda t} \end{aligned}$$

The probability density function of some atom decaying within t time:

$$\begin{aligned} \rho_Y(t) &= \frac{d}{dt} m_Y(t) \\ &= \frac{d}{dt} (1 - e^{-\lambda t}) \\ &= \lambda e^{-\lambda t} \end{aligned}$$

3 Linear Transformations

Let X be a continuous random variable with a known distribution and hence density function. Define Y to be a linear transformation of X . What is the density function of Y ?

Case 1: $Y := aX + b$ where a is positive

$$\begin{aligned} m_Y(y) &= \mathbb{P}(Y \leq y) \\ &= \mathbb{P}(aX + b \leq y) \\ &= \mathbb{P}(X \leq (y - b)/a) \\ &= m_X((y - b)/a) \end{aligned}$$

$$\begin{aligned} \rho_Y(y) &= \frac{d}{dy} m_Y(y) \\ &= \frac{d}{dy} m_X((y - b)/a) \\ &= \rho_X((y - b)/a) \cdot \frac{d}{dy} ((y - b)/a) \\ &= \rho_X((y - b)/a) \cdot \frac{1}{a} \\ &= \frac{1}{a} \rho_X((y - b)/a) \end{aligned}$$

Case 2: $Y := aX + b$ where a is negative

$$\begin{aligned} m_Y(y) &= \mathbb{P}(Y \leq y) \\ &= \mathbb{P}(aX + b \leq y) \\ &= \mathbb{P}(X \geq (y - b)/a) \\ &= 1 - \mathbb{P}(X < (y - b)/a) \\ &= 1 - m_X((y - b)/a) \end{aligned}$$

$$\begin{aligned} \rho_Y(y) &= \frac{d}{dy} m_Y(y) \\ &= \frac{d}{dy} (1 - m_X((y - b)/a)) \\ &= -\rho_X((y - b)/a) \cdot \frac{d}{dy} ((y - b)/a) \\ &= -\rho_X((y - b)/a) \cdot \frac{1}{a} \\ &= -\frac{1}{a} \rho_X((y - b)/a) \end{aligned}$$

We combine these results to get $\rho_Y(y) = \frac{1}{|a|} \rho_X(\frac{y-b}{a})$.

4 Total Weighted Results

4.1 Results

Result functions map a random variable to a real number - the result of the random variable. An example is if the number of points I get after rolling a 20-sided die is 5 times my roll amount. So a roll of 15 earns me 60 points. Such a function would be defined $r(x) = 5x$ where x is the random variable.

4.2 Weighted results

Weighing a result is getting the product of the result and the probability of a random variable. An example is entering a giveaway with a one-in-a-million chance of winning £1,000,000. The weighted result from entering the giveaway would be £1,000,000 · (1/1,000,000) = £1; the average person should expect £1 in value from entering the giveaway.

4.3 Total weighted results

Total weighted results are the sum of all the weighted results of a random variable. An example is buying a £10 lottery ticket with a one-in-a-thousand chance of winning £1000. The total weighted result, accounting for ticket prices, would be $(£1000-£10) \cdot (1/1000) + (£0-£10) \cdot (1-1/1000) = -£9$; the average person should expect to lose £9 when buying the lottery ticket.

4.4 Formulae

Let r be the result function and p be the probability function. Then the total weighted result \mathbb{R} of a discretely distributed x is given by $\sum r(x)p(x)$ for all random variables x .

Let r be the result function and ρ be the probability density function. Then the total weighted result \mathbb{R} of a continuously distributed x is given by $\int_{-\infty}^{\infty} r(x)\rho(x) dx$

4.5 Expectation and Variance

$$\mathbb{E}[X] = \sum x_i \cdot \rho_X(x_i) \quad \text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

$$\mathbb{E}[X] = \int x \cdot \rho_X(x) dx \quad \text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b \quad \text{Var}(aX + b) = a^2\text{Var}(X)$$

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y] \quad \text{Var}(X) = \sigma^2$$

$$\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$$

$$\mathbb{E}[f(X) \cdot g(Y)] = \mathbb{E}[f(X)] \cdot \mathbb{E}[g(Y)]$$

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

4.6 Expectation and Variance of the Bernoulli Distribution

Let $X \sim \text{Bernoulli}(p)$. Then

$$\begin{aligned} \mathbb{E}[X] &= \sum x_i \cdot \rho(x_i) & \text{Var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\ &= 1 \cdot p + 0 \cdot (1 - p) & &= p - p^2 \\ &= p & &= p(1 - p) \end{aligned}$$

4.7 Expectation and Variance of the Binomial Distribution

Let $X \sim \text{Binomial}(n, p)$. Then

$$\begin{aligned} \mathbb{E}[X] &= \mathbb{E}(\sum_{i=0}^n X_i) & \text{Var}(X) &= \text{Var}(\sum_{i=0}^n X_i) \\ &= \sum_{i=0}^n \mathbb{E}[X_i] & &= \sum_{i=0}^n \text{Var}(X_i) \text{ since all } X_i \text{ are independent} \\ &= \sum_{i=0}^n p & &= \sum_{i=0}^n p(1 - p) \\ &= np & &= np(1 - p) \end{aligned}$$

4.8 Expectation of the Geometric Distribution

Let $X \sim \text{Geometric}(p)$. Then

$$\begin{aligned} \mathbb{E}[X] &= \sum_{i=1}^{\infty} x_i \cdot \rho(x_i) \\ \mathbb{E}[X] &= \sum_{i=1}^{\infty} i \cdot (1 - p)^{i-1} p \end{aligned}$$

$$\begin{aligned}
\mathbb{E}[X](1-p) &= \sum_{i=1}^{\infty} i \cdot (1-p)^i p \\
\mathbb{E}[X] - \mathbb{E}[X](1-p) &= \sum_{i=1}^{\infty} i \cdot (1-p)^{i-1} p - \sum_{i=1}^{\infty} i \cdot (1-p)^i p \\
\mathbb{E}[X]p &= \sum_{i=1}^{\infty} i \cdot (1-p)^{i-1} p - \sum_{i=2}^{\infty} (i-1) \cdot (1-p)^{i-1} p \\
\mathbb{E}[X]p &= \sum_{i=1}^{\infty} i \cdot (1-p)^{i-1} p - \sum_{i=2}^{\infty} i \cdot (1-p)^{i-1} p + \sum_{i=2}^{\infty} 1 \cdot (1-p)^{i-1} p \\
\mathbb{E}[X]p &= \sum_{i=1}^1 i \cdot (1-p)^{i-1} p + \sum_{i=2}^{\infty} (1-p)^{i-1} p \\
\mathbb{E}[X]p &= \sum_{i=1}^1 (1-p)^{i-1} p + \sum_{i=2}^{\infty} (1-p)^{i-1} p \\
\mathbb{E}[X]p &= \sum_{i=1}^{\infty} (1-p)^{i-1} p \\
\mathbb{E}[X]p &= \sum_{i=1}^{\infty} \rho(x_i) \\
\mathbb{E}[X]p &= 1 \\
\mathbb{E}[X] &= \frac{1}{p}
\end{aligned}$$

4.9 Expectation of the Poisson Distribution

Let $X \sim \text{Poisson}(\lambda)$. Then

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{m=0}^{\infty} (m \cdot \rho_X(m)) \\
&= \sum_{m=0}^{\infty} \left(m \cdot \frac{\lambda^m}{m!} e^{-\lambda} \right) \\
&= \sum_{m=0}^{\infty} \left(m \cdot \frac{\lambda^m}{m!} e^{-\lambda} \right) + \sum_{m=1}^{\infty} \left(m \cdot \frac{\lambda^m}{m!} e^{-\lambda} \right) \\
&= \left(0 \cdot \frac{\lambda^0}{0!} e^{-\lambda} \right) + \lambda \sum_{m=1}^{\infty} \frac{\lambda^{m-1}}{(m-1)!} e^{-\lambda} \\
&= 0 + \lambda \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} e^{-\lambda} \\
&= \lambda \sum_{m=0}^{\infty} \rho_X(m) \\
&= \lambda \cdot 1 \\
&= \lambda
\end{aligned}$$

5 Normal distribution

5.1 Standardisation

$$\begin{aligned}
Y &:= \frac{X - \mathbb{E}[X]}{\sqrt{\text{Var}(X)}} \\
\mathbb{E}[Y] &= \mathbb{E}\left[\frac{X - \mathbb{E}[X]}{\sqrt{\text{Var}(X)}}\right] \\
&= \frac{\mathbb{E}[X] - \mathbb{E}[X]}{\sqrt{\text{Var}(X)}} \\
&= 0
\end{aligned}$$

$$\begin{aligned}
\text{Var}(Y) &= \text{Var}\left(\frac{X - \mathbb{E}[X]}{\sqrt{\text{Var}(X)}}\right) \\
&= \frac{\text{Var}(X)}{\sqrt{\text{Var}(X)}^2} \\
&= 1
\end{aligned}$$

$$\rho_Y(x) = \sqrt{\text{Var}(X)} \cdot \rho_X(\sqrt{\text{Var}(X)} \cdot x + \mathbb{E}[X])$$

5.2 Normal Distribution Derivation

We have the function e^{-x^2} that we now make into a PDF. To do so its CDF must equal 1 from $-\infty$ to ∞ . So we scale the PDF down to $\frac{1}{\sqrt{\pi}}e^{-x^2}$. What is the expectation and variance of a random variable with this PDF? Somewhere we calculated it to be 0 and $\frac{1}{2}$ respectively. What we want is a PDF of expectation and variance 0 and 1 respectively. We apply the linear transformation lemma to get this PDF: $\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$.

The CDF of the distribution with this PDF cannot be expressed with our usual mathematical operators and functions. Some smart people have evaluated this CDF using other numerical methods at many points already so

we use those predetermined values in evaluating this CDF.

Say we want to model a distribution but our standard normal distribution does not fit it perfectly. We can linearly transform this distribution into one that does and then any CDF values we want to find in the new distribution we get by just performing the same linear transformation on the corresponding predetermined values of the standard normal.

Example: we have a random variable with a pdf that follows a normal distribution of mean $\mu = 10$ and variance $\sigma^2 = 4$. Hence the standard deviation is $\sigma = 2$. Any value x on this normal distribution we standardise by $\frac{x-\mu}{\sigma}$. Now we evaluate the CDF of the standard normal distribution at this point (remember we assumed our distribution is similar to this) by looking up our number on the table.

5.3 Standard Normal Distribution

$X \sim \text{Normal}(0, 1)$

Expectation and Variance:

$$\mathbb{E}[X] = 0 \quad \text{Var}(X) = 1$$

Density and mass:

$$\rho_X(x) := \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad m_X(x) := \int_{-\infty}^x \phi(u) du$$

Define:

$$\phi(x) = \rho_X(x) \quad \Phi(x) = m_X(x)$$

5.4 Gaussian Distribution

$$Y := \sigma X + \mu$$

$$Y \sim \text{Normal}(\mu, \sigma^2)$$

Expectation and variance:

$$\begin{aligned} \mathbb{E}[Y] &= \mathbb{E}[\sigma X + \mu] & \text{Var}(Y) &= \text{Var}(\sigma X + \mu) \\ &= \sigma \mathbb{E}[X] + \mu & &= \sigma^2 \text{Var}(X) \\ &= \sigma \cdot 0 + \mu & &= \sigma^2 \cdot 1 \\ &= \mu & &= \sigma^2 \end{aligned}$$

Density and mass:

$$\begin{aligned} \rho_Y(y) &= \frac{1}{\sigma} \rho_X\left(\frac{y-\mu}{\sigma}\right) & m_Y(y) &= m_X\left(\frac{y-\mu}{\sigma}\right) \\ &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}} & &= \text{Phi}\left(\frac{y-\mu}{\sigma}\right) \end{aligned}$$

5.5 Central Limit Theorem

$$\lim_{n \rightarrow \infty} \frac{X_n - \mathbb{E}[X_n]}{\sqrt{\text{Var}(X_n)}} = \text{Normal}(0, 1)$$

6 Multiple Random Variables

6.1 Joint Probabilities for Discrete Random Variables

Let X and Y be two discrete random variables.

The joint probability functions of X and Y are given by

- $p_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y)$;
- $\mathbb{P}(X \in A, Y \in B) = \sum_{x \in A} \sum_{y \in B} p_{X,Y}(x, y)$.

The marginal probability functions of X and Y are

- $p_X(x) = \sum_y p_{X,Y}(x, y)$;
- $p_Y(y) = \sum_x p_{X,Y}(x, y)$.

The joint probability functions follow the rules as normal:

- $0 \leq p_{X,Y}(x, y) \leq 1$ for all x and y ;
- $\sum_x \sum_y p_{X,Y}(x, y) = 1$.

X and Y are independent if and only if $p_{X,Y}(x, y) = p_X(x) \cdot p_Y(y)$.

6.2 Joint Densities for Continuous Random Variables

Let X and Y be two continuous random variables.

The joint density function of X and Y is given by

- $\rho_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y)$;
- $m_{X,Y}(x, y) = \mathbb{P}(X \in A, Y \in B)$;
- $\mathbb{P}(X \in A, Y \in B) = \int_A \int_B \rho_{X,Y}(x, y) dy dx$.

The marginal density functions of X and Y are given by

- $\rho_X(x) = \int_B \rho_{X,Y}(x, y) dy$;
- $\rho_Y(y) = \int_A \rho_{X,Y}(x, y) dx$.

The joint density function follows the rules:

- $\rho_{X,Y}(x, y) \geq 0$;
- $\int_{\mathbb{R}} \int_{\mathbb{R}} \rho_{X,Y}(x, y) dy dx = 1$.

X and Y are independent if and only if $\rho_{X,Y}(x, y) = \rho_X(x) \cdot \rho_Y(y)$.

6.3 Expectations for Joint Probabilities and Densities

Let X and Y be discrete random variables. Then $\mathbb{E}[f(X, Y)] = \sum_x \sum_y f(X, Y) \cdot p_{X,Y}(x, y)$.

It follows that $\mathbb{E}[f(X)] = \sum_x \sum_y f(X) \cdot p_{X,Y}(x, y) = \sum_x f(X) \cdot \sum_y p_{X,Y}(x, y) = \sum_x f(X) \cdot p_X(x)$.

Let X and Y be continuous random variables. Then $\mathbb{E}[f(X, Y)] = \int_{\mathbb{R}} \int_{\mathbb{R}} f(X, Y) \cdot \rho_{X,Y}(x, y) dy dx$.

It follows that $\mathbb{E}[f(X)] = \int_{\mathbb{R}} \int_{\mathbb{R}} f(X) \cdot \rho_{X,Y}(x, y) dy dx = \int_{\mathbb{R}} f(X) \cdot \int_{\mathbb{R}} \rho_{X,Y}(x, y) dy dx = \int_{\mathbb{R}} f(X) \cdot \rho_X(x) dx$.

If X and Y are independent then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$

6.4 Covariance and Correlation

Let X and Y be random variables.

Define **covariance** as a measure of how dependent X and Y are on each other:

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

Covariance has properties:

- $\text{Cov}(X, X) = \text{Var}(X)$;
- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$;
- $\text{Cov}(aX, Y) = a\text{Cov}(X, Y)$;
- $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$;

Covariance has the equation $\text{Cov}(X, Y) = 0$ if X and Y are independent.

Define **correlation** as a standardised covariance: $\text{Cor}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$

Correlation has the properties:

- $-1 \leq \text{Cor}(X, Y) \leq 1$
- $\text{Cov}(X, Y)^2 \leq \text{Var}(X)\text{Var}(Y)$

Correlation has the equations:

- $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$;
- $|\text{Cor}(X, Y)| = 1$ if and only if Y is a linear transformation of X .

7 Functions of Random Variables

7.1 Maximum and Minimum of Random Variables

Let $X_{\max} = \max\{X_1, X_2, \dots, X_n\}$ and $X_{\min} = \min\{X_1, X_2, \dots, X_n\}$ where X_i are independent and identically distributed random variables. Then

$$\begin{aligned}
 m_{X_{\max}}(x) &= \mathbb{P}(X_{\max} \leq x) \\
 &= \mathbb{P}(X_1 \leq x, X_2 \leq x, \dots, X_n \leq x) \\
 &= \mathbb{P}(X_1 \leq x) \cdot \mathbb{P}(X_2 \leq x) \cdot \dots \cdot \mathbb{P}(X_n \leq x) \\
 &= \mathbb{P}(X \leq x)^n \text{ since they are all identical} \\
 &= m_X(x)^n;
 \end{aligned}
 \qquad
 \begin{aligned}
 m_{X_{\min}}(x) &= \mathbb{P}(X_{\min} \leq x) \\
 &= 1 - \mathbb{P}(X_{\min} > x) \\
 &= 1 - \mathbb{P}(X_1 > x, X_2 > x, \dots, X_n > x) \\
 &= 1 - \mathbb{P}(X_1 > x) \cdot \mathbb{P}(X_2 > x) \cdot \dots \cdot \mathbb{P}(X_n > x) \\
 &= 1 - \mathbb{P}(X > x)^n \text{ since they are all identical} \\
 &= 1 - (1 - \mathbb{P}(X \leq x))^n \\
 &= 1 - (1 - m_X(x))^n.
 \end{aligned}$$

7.2 Sums of Random Variables

Let $Z = X + Y$ where X and Y are independent and discrete random variables. Then

$$p_Z(z) = \sum_y p_X(z - y)p_Y(y) \quad \text{or} \quad p_Z(z) = \sum_x p_X(x)p_Y(z - x).$$

Let $Z = X + Y$ where X and Y are independent and continuous random variables. Then

$$m_Z(z) = \int_{\mathbb{R}} \rho_X(z - y)\rho_Y(y) dy \quad \text{or} \quad m_Z(z) = \int_{\mathbb{R}} \rho_X(x)\rho_Y(z - x) dx.$$

7.3 Sums of Random Variables of Known Distributions

$$\begin{aligned}
X \sim \text{Bernoulli}(p) & \quad \& \quad Y \sim \text{Bernoulli}(p) & \quad \implies & \quad X + Y \sim \text{Binomial}(2, p). \\
X \sim \text{Binomial}(n_1, p) & \quad \& \quad Y \sim \text{Binomial}(n_2, p) & \quad \implies & \quad X + Y \sim \text{Binomial}(n_1 + n_2, p). \\
X \sim \text{Poisson}(\lambda_1) & \quad \& \quad Y \sim \text{Poisson}(\lambda_2) & \quad \implies & \quad X + Y \sim \text{Poisson}(\lambda_1 + \lambda_2). \\
X \sim \text{Exponential}(\lambda_1) & \quad \& \quad Y \sim \text{Exponential}(\lambda_2) & \quad \implies & \quad \min X, Y \sim \text{Exponential}(\lambda_1 + \lambda_2). \\
X \sim \text{Exponential}(\lambda) & \quad \& \quad Y \sim \text{Exponential}(\lambda) & \quad \implies & \quad X + Y \sim \text{Gamma}(2, \lambda). \\
X \sim \text{Gamma}(\alpha, \lambda) & \quad \& \quad Y \sim \text{Gamma}(\beta, \lambda) & \quad \implies & \quad X + Y \sim \text{Gamma}(\alpha + \beta, \lambda). \\
X \sim \text{Gaussian}(\mu_1, \sigma_1^2) & \quad \& \quad Y \sim \text{Gaussian}(\mu_2, \sigma_2^2) & \quad \implies & \quad X + Y \sim \text{Gaussian}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).
\end{aligned}$$

8 More stuff

8.1 Conditional Probability

Definition: Let X be a discrete random variable and A be a probable event. Conditional expectation is given by:

$$\mathbb{E}[X|A] = \sum_i x_i \mathbb{P}(X = x_i|A).$$

Identity: Let $\Omega = A_1 \sqcup A_2 \sqcup \dots \sqcup A_n$. Conditional expectation follows the law of total probability:

$$\mathbb{P}(X = x) = \mathbb{P}(X = x|A_1)\mathbb{P}(A_1) + \mathbb{P}(X = x|A_2)\mathbb{P}(A_2) + \dots + \mathbb{P}(X = x|A_n)\mathbb{P}(A_n).$$

Theorem: Let A be the event that a positive integer n is divisible by a , B be the event that n is divisible by b . A and B are independent if and only if $\gcd(a, b) = 1$.

- n is divisible by a and b if and only if n is divisible by $\text{lcm}(a, b) = \frac{ab}{\gcd(a, b)}$;
- $\mathbb{P}(A) = \frac{1}{a}$ and $\mathbb{P}(B) = \frac{1}{b}$ and $\mathbb{P}(A \cap B) = \frac{1}{\text{lcm}(a, b)} = \frac{\gcd(a, b)}{ab}$;
- $\mathbb{P}(A) \cdot \mathbb{P}(B) = \mathbb{P}(A \cap B)$ if and only if $\gcd(a, b) = 1$.

Example: Let A be the probability that a positive integer n is divisible by 2 and B be the probability that n is divisible by 3. Find the probability that n is divisible by 2 or 3.

$$\begin{aligned}
\mathbb{P}(A \cup B) &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \quad (\text{the inclusion-exclusion principle}) \\
&= \frac{1}{2} + \frac{1}{3} - \frac{1}{2 \cdot 3} \\
&= \frac{2}{3}.
\end{aligned}$$

8.2 Cheyshev's Inequality and the LLN

Cheyshev's Inequality: Let X be a random variable and d be any positive number. The probability of the distance between X and $\mathbb{E}[X]$ being greater than a is bounded:

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq d) \leq \frac{\text{Var}(X)}{d^2}.$$

Weak Law of Large Numbers: Let X_i be independent and identically distributed random variables and d be any positive number. The probability of the distance between \bar{X} and $\mathbb{E}[X]$ being greater than d tends to 0:

$$\lim_{n \rightarrow \infty} \mathbb{P}(|\frac{1}{n} \sum_{i=1}^n X_i - \mathbb{E}[X]| \geq d) = 0$$

Example: Find the expected number of consecutive heads if you toss a coin n times.

- Let X_i equal 1 if the consecutive i th and $i + 1$ th tosses are both heads and 0 otherwise.
- The expected number of consecutive heads in n tosses is simply $\mathbb{E}[X_1 + X_2 + \dots + X_{n-1}]$.

$$\begin{aligned}
 & \mathbb{E}[X_1 + X_2 + \dots + X_{n-1}] \\
 &= \mathbb{E}[\sum_{i=1}^{n-1} X_i] \\
 &= \sum_{i=1}^{n-1} \mathbb{E}[X_i] \text{ (linearity of expectation)} \\
 &= \sum_{i=1}^{n-1} \mathbb{E}[X] \text{ (probability of consecutive heads is all equal)} \\
 &= \sum_{i=1}^{n-1} \frac{1}{4} \\
 &= \frac{n-1}{4}
 \end{aligned}$$

Example: Find a value for n where n is number of coin tosses needed for the probability of the proportion of heads to be within 0.01 of 0.5 to be least 95%. Let $X \sim \text{Binomial}(n, 0.5)$ be the number of heads and $Y = \frac{1}{n}X$ be the proportion of heads.

$$\begin{aligned}
 \mathbb{E}[Y] &= \mathbb{E}[\frac{1}{n}X] & \text{Var}(Y) &= \text{Var}(\frac{1}{n} \cdot X) \\
 &= \frac{1}{n} \mathbb{E}[X] & &= \frac{1}{n^2} \cdot \text{Var}(X) \\
 &= \frac{1}{n} np & &= \frac{1}{n^2} \cdot np(1-p) \\
 &= \frac{1}{n} \sum_{i=1}^n \frac{n}{2} & &= \frac{1}{n^2} \cdot n \cdot \frac{1}{2} (1 - \frac{1}{2}) \\
 &= \frac{1}{2} & &= \frac{1}{4n}
 \end{aligned}$$

$$\begin{aligned}
 & \mathbb{P}(|X - E[X]| \geq d) \leq \frac{\text{Var}(X)}{d^2} \\
 & \mathbb{P}(|X - 0.5| \geq 0.01) \leq \frac{2500}{n} \\
 & 1 - \frac{2500}{n} \leq \mathbb{P}(|X - 0.5| < 0.01) \\
 & 0.95 \leq 1 - \frac{2500}{n} \leq \mathbb{P}(|X - 0.5| < 0.01) \quad \text{we can't calculate } \mathbb{P} \text{ directly so just do this transitive trick} \\
 & 0.95 \leq 1 - \frac{2500}{n} \quad \text{find a } n \text{ such that } 0.95 \leq \mathbb{P}(|X - 0.5| < 0.01) \\
 & n \geq 50000
 \end{aligned}$$