

Mathematics/Statistics Bootcamp

Part IV: Probability

Steven Winter Christine Shen

Department of Statistical Science
Duke University

MSS Orientation, August 2022

Outline

Probability

- Independence

- Bayes' Rule

Multivariate Distributions

- Joint Distribution

- Marginal Distribution

Moments

- Expectation, Variance and Covariance

- Kernel Trick

- Moment Generating Functions

Probability

Axioms of Probability

1. For any event A , $\mathbb{P}(A) \in [0, 1]$;
2. Let Ω denote the sample space, $\mathbb{P}(\Omega) = 1$;
3. If A_1, A_2, \dots are disjoint events, then

$$\mathbb{P}\left(\bigcup_i A_i\right) = \sum_{i=1} \mathbb{P}(A_i).$$

Independence

Consider two events A and B in the sample space Ω , $\mathbb{P}(B) > 0$. We say A and B are **independent** (denoted $A \perp B$) if:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

A collection of events A_1, \dots, A_n are considered **mutually independent** if for *any* sub-collection A_{i_1}, \dots, A_{i_K} we have:

$$\mathbb{P}(\cap_{j=1}^K A_{i_j}) = \prod_{j=1}^K \mathbb{P}(A_{i_j}).$$

Independence - Example

Consider an experiment of tossing two dice. The sample space is

$$\Omega = \{(1, 1), (1, 2), \dots, (1, 6), (2, 1), \dots, (2, 6), \dots, (6, 6)\}.$$

Further consider the following events:

$$A = \{\text{doubles appear}\}$$

$$B = \{\text{the sum is between 7 and 10}\}$$

$$C = \{\text{the sum is 2 or 7 or 10}\}.$$

Are A, B, C mutually independent?

Discussion

Let A , B and C be events.

1. If $A \perp A$, what do we know about A ?
2. If $A \perp B$, is $A \perp B^c$?
3. If $A \perp B$, and $B \perp C$, is $A \perp C$?

Conditional Probability

Consider two events A and B in the sample space Ω , $\mathbb{P}(B) > 0$. The **conditional probability** of event A given B is defined as

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Let A , B and C be events. A and B are said to be **conditionally independent** given C if and only if $\mathbb{P}(C) > 0$, and

$$\mathbb{P}(A \cap B \mid C) = \mathbb{P}(A \mid C)\mathbb{P}(B \mid C).$$

Usually we denote as $A \perp B \mid C$.

Law of Total Probability and Bayes Rule

A countable collection of events $\{A_1, A_2, \dots\}$ is called a partition if $A_i \cap A_j = \emptyset$ for $i \neq j$, and $\cup_j A_j = \Omega$. Let B be an event such that $\mathbb{P}(B) > 0$.

Law of Total Probability:

$$\mathbb{P}(B) = \sum_j \mathbb{P}(A_j) \mathbb{P}(B \mid A_j).$$

Bayes Rule:

$$\mathbb{P}(A_i \mid B) = \frac{\mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A_i) \mathbb{P}(B \mid A_i)}{\mathbb{P}(B)}.$$

Putting them together, we have:

$$\mathbb{P}(A_i \mid B) = \frac{\mathbb{P}(A_i) \mathbb{P}(B \mid A_i)}{\sum_j \mathbb{P}(A_j) \mathbb{P}(B \mid A_j)}.$$

Example

In Morse code, information is represented as dots and dashes.
Assume the following:

$$\mathbb{P}(\textit{dot sent}) = \frac{3}{7}$$

$$\mathbb{P}(\textit{dash sent}) = \frac{4}{7}$$

$$\mathbb{P}(\textit{dot received}|\textit{dot sent}) = \frac{7}{8}$$

Find $\mathbb{P}(\textit{dot sent}|\textit{dot received})$.

Exercises

1. Consider all length 3 strings constructable from $\{a, b, c\}$:

$$\Omega = \{aaa, bbb, ccc, abc, bca, cba, acb, bac, cab\}.$$

Assign each string probability $\frac{1}{9}$. For $i = 1, 2, 3$, define A_i as:

$$A_i = \{i^{th} \text{ place in the triple is occupied by } a\}.$$

Are the A_i independent? Prove/disprove.

2. Assume we know the following about a specific disease, D :
 - ▶ the probability of getting sick is 0.01
 - ▶ the probability of testing positive if sick is 0.95
 - ▶ the probability of testing negative if healthy is 0.95

What is the probability of being sick if the test is positive?

Multivariate Distributions

Joint Distribution

Joint PDF: A function $f(x_1, \dots, x_n)$ from $\mathbb{R}^n \rightarrow \mathbb{R}$ is called a joint PDF of the random vector $\mathbf{X} = (X_1, \dots, X_n)$ if for every $A \subset \mathbb{R}^n$,

$$\mathbb{P}(\mathbf{X} \in A) = \int_A f_{X_1, \dots, X_n}(x_1, \dots, x_n) d(x_1, \dots, x_n).$$

Joint PMF: Let R_{X_i} denote the range of discrete variable X_i , $R_{\mathbf{X}} = R_{X_1} \times \dots \times R_{X_n}$. Let

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = \mathbb{P}(X_1 = x_1, \dots, X_n = x_n)$$

be the joint PMF of $\mathbf{X} = (X_1, \dots, X_n)$. Then for every $A \subset \mathbb{R}^n$,

$$\mathbb{P}(\mathbf{X} \in A) = \sum_{(x_1, \dots, x_n) \in (A \cap R_{\mathbf{X}})} f_{X_1, \dots, X_n}(x_1, \dots, x_n).$$

Marginal Distribution

Given the joint PDF/ PMF, we can find the marginal PDF/ PMF:

Marginal PDF:

$$f_{X_1}(x_1) = \int_{X_2, \dots, X_n} f_{X_1, \dots, X_n}(x_1, \dots, x_n) d(x_2 \dots x_n).$$

Marginal PMF:

$$f_{X_1}(x_1) = \sum_{(x_2, \dots, x_n) \in (R_{X_2} \times \dots \times R_{X_n})} f_{X_1, \dots, X_n}(x_1, \dots, x_n).$$

Joint Distribution - Exercise

1. Assume that X and Y have the joint PDF:

$$f_{X,Y}(x,y) = 4xy, \quad 0 < x < 1 \quad 0 < y < 1.$$

Find $\mathbb{P}(Y < X)$.

2. Random variables X and Y are jointly normal with mean $(\mu_x, \mu_y)^T$ and covariance matrix

$$\begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}.$$

Find $\mathbb{P}(Y < X)$. Think about what happens if $\mu_x \rightarrow \infty$?
What about limiting cases of other parameters?

Hint:

- ▶ What's the distribution of $Y - X$?
- ▶ $\mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y] + 2\text{Cov}(X, Y)$.

Conditional Distribution

Let X, Y be random variables with joint PDF/ PMF $f_{X,Y}(x, y)$.
The **conditional PDF/ PMF** of X given $Y = y$ is:

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

For discrete random variables, this is intuitive:

$$f_{X|Y}(x | y) = \mathbb{P}(X = x | Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)} = \frac{f_{X,Y}(x, y)}{f_Y(y)}.$$

How to understand it for continuous random variables? What's $\mathbb{P}(Y = y)$ for a continuous random variable Y ?

Conditional Distribution

For continuous random variables X and Y , $\mathbb{P}(Y = y) = 0$, and

$$f_{X|Y}(x | y) \neq \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}.$$

Instead, $f_{X|Y}(x | y)$ is the function such that

$$\int_B f_{X|Y}(x | y) dx = \mathbb{P}(X \in B | Y = y), \quad \text{and}$$
$$\mathbb{P}(X \in B | Y = y) = \lim_{\epsilon \rightarrow 0} \mathbb{P}(X \in B | Y \in (y, y + \epsilon)).$$

Conditional Distribution - Exercise

1. Assume that (X, Y) is a continuous random vector with joint pdf given by:

$$f_{X,Y}(x, y) = e^{-y}, \quad 0 < x < y < \infty.$$

Find the marginal distribution of X , and the conditional distribution $Y|X$.

2. Let $Y \sim N(\mu, \sigma^2)$ with known μ and σ^2 . Find the PDF for $Y \mid Y \geq c$, for some $c \in \mathbb{R}$.

Bonus: Generalize this to a standard multi-variate normal, $\mathbf{Z} \sim N_n(\mathbf{0}, \mathbf{I})$, by finding the PDF for $\mathbf{Z} \mid \mathbf{Z} \in \mathbb{R}_+^n$. What happens in high dimensions (when $n \rightarrow \infty$)?

Conditional Independence

Let A , B and C be events. Recall A and B are said to be **conditionally independent** given C if and only if $\mathbb{P}(C) > 0$, and

$$\mathbb{P}(A \cap B \mid C) = \mathbb{P}(A \mid C)\mathbb{P}(B \mid C).$$

Usually we denote as $A \perp B \mid C$.

An equivalent definition is

$$\mathbb{P}(A \mid B, C) = \mathbb{P}(A \mid C).$$

Self exercise: prove these two definitions are equivalent!

Conditional Independence

Similarly, random variables X and Y are **conditionally independent** given random variable Z if and only if

$$f_{X,Y|Z=z}(x,y) = f_{X|Z=z}(x)f_{Y|Z=z}(y),$$

where $f_{\cdot|Z}(\cdot)$ is the conditional PDF/ PMF given Z .

Usually we denote as $X \perp Y \mid Z$.

Conditional Independence - Example

Suppose we have three discrete random variables Y_1, Y_2, Y_3 that we believe are "independent and identically distributed (i.i.d.)". Does our knowledge about the value of one inform about another? That is:

$$\mathbb{P}(Y_1 = y_1 \mid Y_2 = y_2, Y_3 = y_3) = \mathbb{P}(Y_1 = y_1)?$$

What if Y_1, Y_2, Y_3 are conditionally independent given discrete random variable Θ ?

Moments

Expectation of Random Variables

Let X be an integrable random variable,¹ $f_X(x)$ be its PDF/PMF, and $g : \mathbb{R} \rightarrow \mathbb{R}$ be any real function. The expectation of $g(X)$ is:

- ▶ if X is continuous,

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x)dx.$$

- ▶ if X is discrete, let \mathcal{X} denote its range,

$$\mathbb{E}[g(X)] = \sum_{x \in \mathcal{X}} g(x)f_X(x) = \sum_{x \in \mathcal{X}} g(x)\mathbb{P}(X = x).$$

Setting $g(X) = X$ gives $\mathbb{E}[X]$, the expectation of X .

¹i.e., expectation of X exists. Counter-example: expectation of a Cauchy random variable is undefined.

Variance and Covariance of Random Variables

Let X, Y be square integrable random variables.² Variance of X is defined as

$$\begin{aligned}\mathbb{V}[X] &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\ &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2.\end{aligned}$$

Covariance between X and Y is defined as

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

²i.e., both expectation and variance exist

Expectation and Variance - Exercise

$X \sim \text{Poisson}(\lambda)$. Show that $\mathbb{E}[X] = \lambda$.

Properties of Expectation

Let

- ▶ X, Y be integrable random variables
- ▶ $a \in \mathbb{R}$ be a scalar constant
- ▶ f and $g : \mathbb{R} \rightarrow \mathbb{R}$ be functions such that $f(X)$ and $g(X)$ are integrable

Basic properties of Expectation:

1. Linearity

- ▶ $\mathbb{E}[aX] = a\mathbb{E}[X]$
- ▶ $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$

2. Monotonicity

- ▶ $f \leq g \implies \mathbb{E}[f(X)] \leq \mathbb{E}[g(X)]$, or equivalently,
- ▶ $X \leq Y$ with probability 1 $\implies \mathbb{E}[X] \leq \mathbb{E}[Y]$

Jensen's Inequality

Convex function

- ▶ A function $\psi : \mathcal{X} \rightarrow \mathbb{R}$ is convex iff for all $t \in [0, 1]$, $x_1, x_2 \in \mathcal{X}$,

$$f(tx_1 + (1 - t)x_2) \leq tf(x_1) + (1 - t)f(x_2).$$

It is strictly convex if for any $x_1 \neq x_2$, the inequality is strict.

- ▶ Any twice differentiable function ψ is convex iff its second derivative is non-negative. It is strictly convex if its second derivative is positive.

By **Jensen's inequality**, for any integrable random variable X , and convex function ψ ,

$$\psi(\mathbb{E}[X]) \leq \mathbb{E}[\psi(X)].$$

Inequality is strict if ψ is strictly convex and X is non-degenerate.

Jensen's Inequality - Optional Example

Let $\|X\|_p = \mathbb{E}[X^p]^{1/p}$ denote the L_p norm of a random variable X .

For $0 < p < q < \infty$, let X be a random variable such that X^q is integrable. Use Jensen's inequality to show

$$\|X\|_p \leq \|X\|_q.$$

Cauchy-Schwartz and Hölder's Inequalities

Cauchy-Schwartz inequality

For any square integrable random variables X and Y ,

$$\mathbb{E}[XY] \leq \mathbb{E}[|XY|] \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}.$$

Cauchy-Schwartz is a special case of **Hölder's inequality**

For $r \geq 1$, $p, q > 1$ with $1/p + 1/q = 1/r$,

$$\|XY\|_r \leq \|X\|_p \|Y\|_q.$$

Expectation - Example

1. Let \mathbf{A} be an $n \times n$ random matrix, show

$$\mathbb{E}[\text{Tr}(\mathbf{A})] = \text{Tr}(\mathbb{E}[\mathbf{A}]).$$

Expectation - Example

1. Let \mathbf{A} be an $n \times n$ random matrix, show

$$\mathbb{E}[\text{Tr}(\mathbf{A})] = \text{Tr}(\mathbb{E}[\mathbf{A}]).$$

Proof:

$$\begin{aligned}\mathbb{E}[\text{Tr}(\mathbf{A})] &= \mathbb{E}\left[\sum_{i=1}^n a_{ii}\right] = \sum_{i=1}^n \mathbb{E}[a_{ii}] \\ &= \text{Tr}\left(\begin{pmatrix} \mathbb{E}[a_{11}] & \cdots & \mathbb{E}[a_{1n}] \\ \vdots & \ddots & \vdots \\ \mathbb{E}[a_{n1}] & \cdots & \mathbb{E}[a_{nn}] \end{pmatrix}\right) \\ &= \text{Tr}(\mathbb{E}[\mathbf{A}]).\end{aligned}$$

Expectation - Example Cont.

2. Consider a random vector $\mathbf{Y} \in \mathbb{R}^n$ with $\mathbb{E}[\mathbf{Y}] = \boldsymbol{\mu}$, and $\mathbb{V}[\mathbf{Y}] = \boldsymbol{\Sigma}$. Show that for any fixed matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$,

$$\mathbb{E}[\mathbf{Y}^T \mathbf{A} \mathbf{Y}] = \boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu} + \text{Tr}(\mathbf{A} \boldsymbol{\Sigma}).$$

Expectation - Example Cont.

2. Consider a random vector $\mathbf{Y} \in \mathbb{R}^n$ with $\mathbb{E}[\mathbf{Y}] = \boldsymbol{\mu}$, and $\mathbb{V}[\mathbf{Y}] = \boldsymbol{\Sigma}$. Show that for any fixed matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$,

$$\mathbb{E}[\mathbf{Y}^T \mathbf{A} \mathbf{Y}] = \boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu} + \text{Tr}(\mathbf{A} \boldsymbol{\Sigma}).$$

Proof: Notice

$$\begin{aligned} \mathbf{Y}^T \mathbf{A} \mathbf{Y} &= [\boldsymbol{\mu} + (\mathbf{Y} - \boldsymbol{\mu})]^T \mathbf{A} [\boldsymbol{\mu} + (\mathbf{Y} - \boldsymbol{\mu})] \\ &= \boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu} + (\mathbf{Y} - \boldsymbol{\mu})^T \mathbf{A} \boldsymbol{\mu} + \boldsymbol{\mu}^T \mathbf{A} (\mathbf{Y} - \boldsymbol{\mu}) \\ &\quad + (\mathbf{Y} - \boldsymbol{\mu})^T \mathbf{A} (\mathbf{Y} - \boldsymbol{\mu}). \end{aligned}$$

Taking expectation on both sides, the first term on the RHS is a constant, the middle two terms become zero. For the last term, we can apply the trace trick.

Expectation - Example Cont.

2. Notice that $(\mathbf{Y} - \boldsymbol{\mu})^T \mathbf{A}(\mathbf{Y} - \boldsymbol{\mu})$ is a scalar, therefore

$$\begin{aligned} & \mathbb{E}[(\mathbf{Y} - \boldsymbol{\mu})^T \mathbf{A}(\mathbf{Y} - \boldsymbol{\mu})] \\ &= \mathbb{E}[\text{Tr}[(\mathbf{Y} - \boldsymbol{\mu})^T \mathbf{A}(\mathbf{Y} - \boldsymbol{\mu})]] \\ &= \mathbb{E}[\text{Tr}[\mathbf{A}(\mathbf{Y} - \boldsymbol{\mu})(\mathbf{Y} - \boldsymbol{\mu})^T]] \\ &= \text{Tr}[\mathbb{E}[\mathbf{A}(\mathbf{Y} - \boldsymbol{\mu})(\mathbf{Y} - \boldsymbol{\mu})^T]] \\ &= \text{Tr}[\mathbf{A} \mathbb{E}[(\mathbf{Y} - \boldsymbol{\mu})(\mathbf{Y} - \boldsymbol{\mu})^T]] \\ &= \text{Tr}[\mathbf{A} \boldsymbol{\Sigma}]. \end{aligned}$$

Together with previous results, we have

$$\mathbb{E}[\mathbf{Y}^T \mathbf{A} \mathbf{Y}] = \boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu} + \text{Tr}(\mathbf{A} \boldsymbol{\Sigma}).$$

Properties of Variance

Let X, Y be square integrable random variables, $a, b \in \mathbb{R}$ be scalar constants.

Basic properties of Variance:

1. $\mathbb{V}[X] \geq 0$
2. $\mathbb{V}[X + a] = \mathbb{V}[X]$
3. $\mathbb{V}[aX] = a^2\mathbb{V}[X]$
4. $\mathbb{V}[aX \mp bY] = a^2\mathbb{V}[X] + b^2\mathbb{V}[Y] \mp 2ab\text{Cov}(X, Y)$

Properties of Covariance

Let X, Y, W, V be square integrable random variables,
 $a, b, c, d \in \mathbb{R}$ be scalar constants.

Basic properties of Covariance:

1. $\text{Cov}(X, a) = 0$
2. $\text{Cov}(X, X) = \mathbb{V}[X]$
3. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
4. Bilinearity

$$\text{Cov}(aX + bY, cW + dV) = ac\text{Cov}(X, W) + ad\text{Cov}(X, V) + bc\text{Cov}(Y, W) + bd\text{Cov}(Y, V)$$

Expectation, Variance and Covariance - Example

Assume

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N_2 \left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} \right).$$

We know that the conditional distribution of $X \mid Y$ is also normal.
Find its mean and variance.

Laws of Total Expectation and Total Variance

Let X, Y be square integrable random variables.

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

$$\mathbb{V}[Y] = \mathbb{V}[\mathbb{E}[Y|X]] + \mathbb{E}[\mathbb{V}[Y|X]]$$

Laws of Total Expectation and Total Variance - Example

Consider

$$X|N \sim \text{Binomial}(N, p)$$

$$N \sim \text{Negative Binomial}(\tau, r).$$

Find $\mathbb{E}[X]$ and $\mathbb{V}[X]$.

Hint:

$$\mathbb{E}[N] = \frac{\tau r}{1 - \tau}, \quad \mathbb{V}[N] = \frac{\tau r}{(1 - \tau)^2}.$$

Laws of Total Expectation and Total Variance - Exercise

Consider

$$\begin{aligned}X|P &\sim \text{Binomial}(n, P) \\ P &\sim \text{Beta}(a, b).\end{aligned}$$

Find $\mathbb{E}[X]$ and $\mathbb{V}[X]$.

Hint:

$$\begin{aligned}\mathbb{E}[P] &= \frac{a}{a+b} \\ \mathbb{V}[P] &= \frac{ab}{(a+b)^2(a+b+1)}.\end{aligned}$$

Kernel Trick - Example

Consider $X \sim \text{Exponential}(\lambda)$, with PDF $f_X(x) = \lambda e^{-\lambda x}$.

Moments calculation, e.g., the expectation

$$\mathbb{E}[X] = \int_0^{\infty} x \lambda e^{-\lambda x} dx.$$

usually requires integration by parts.

Kernel Trick - Example Cont.

Alternatively, we can use the **kernel trick** to avoid the tedious calculus.

First, notice that the PDF for $X \sim \text{Gamma}(\alpha, \beta)$ is

$$g_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}.$$

Recall the integral from the previous slide:

$$\mathbb{E}[X] = \int_0^\infty \lambda x e^{-\lambda x} dx.$$

Here the integrand is almost like a Gamma PDF with $\alpha = 2$, $\beta = \lambda$.

Kernel Trick - Example Cont.

The PDF of a random variable integrates to 1. Therefore if we consider $X \sim \text{Gamma}(2, \lambda)$, we have

$$\int_0^{\infty} \frac{\lambda^2}{\Gamma(2)} x e^{-\lambda x} dx = 1.$$

Therefore

$$\begin{aligned}\mathbb{E}[X] &= \int_0^{\infty} \lambda x e^{-\lambda x} dx \\ &= \frac{1}{\lambda/\Gamma(2)} = \frac{1}{\lambda}.\end{aligned}$$

Kernel Trick

The **kernel** of a distribution is the form of the PDF/PMF in which any factors that are not functions of any of the random variable(s) are omitted.

The **kernel trick** utilizes the fact that PDF/PMF integrates/ sums to 1, to help us:

1. solve integration problems (as shown in the last example);
2. identify distributions (see optional exercise in next slide, and also later in Bayesian inference).

Note that the term *kernel* here is different from the *kernel functions* in machine learning.

Kernel Trick - Exercise

Still let $X \sim \text{Exponential}(\lambda)$, use the kernel trick to find $\mathbb{V}[X]$.

Moment Generating Functions

The **moment generating function** (MGF) for a random variable X (if it exists) is defined as:

$$M_x(t) = \mathbb{E}[e^{tX}].$$

Let \mathcal{X} denote the range of X , $f_X(x)$ denote the PDF/ PMF.

- ▶ If X is discrete

$$M_x(t) = \sum_{x \in \mathcal{X}} e^{tx} f_X(x).$$

- ▶ If X is continuous

$$M_x(t) = \int_{\mathcal{X}} e^{tx} f_X(x) dx.$$

Properties of MGF

Let X, Y be random variables with well defined MGFs.

1. If $M_X(t) = M_Y(t)$, then $X \stackrel{d}{=} Y$, i.e., MGF *uniquely* defines the distribution of a random variable.

Exercise: anything else you have learned that can uniquely characterize a distribution?

2. To calculate the n^{th} moment of X

$$\mathbb{E}[X^n] = M_X^{(n)}(0).$$

3. If X and Y are independent,

$$\begin{aligned} M_{X+Y}(t) &= \mathbb{E}[e^{t(X+Y)}] \\ &= \mathbb{E}[e^{tX}] \mathbb{E}[e^{tY}] \\ &= M_X(t) M_Y(t). \end{aligned}$$

MGFs are helpful for determining distributions of sums of independent random variables.

MGF - Example

Let $X \sim \text{Gamma}(\alpha, \beta)$ (rate parameterization). Find $M_X(t)$.

MGF - Exercise

1. Let $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} \text{Gamma}(\alpha, \beta)$, $Y = \sum_{i=1}^n X_i$.

Find $M_Y(t)$, and identify the distribution of Y .

2. (Optional) Let $X_1, \dots, X_N \stackrel{i.i.d.}{\sim} \text{Exponential}(\beta)$, $N \sim \text{Poisson}(\lambda)$, and $Y = \sum_{i=1}^N X_i$. Find $M_Y(t)$.

Hint:

- ▶ $\text{Exponential}(\beta) \stackrel{d}{=} \text{Gamma}(1, \beta)$.
- ▶ Recall the law of total expectation.

Acknowledgement

Past contributors:

- ▶ Jordan Bryan, PhD student
- ▶ Brian Cozzi, MSS alumni
- ▶ Michael Valancius, MSS alumni
- ▶ Graham Tierney, PhD student
- ▶ Becky Tang, PhD student