MATH4240 Tutorial 1

2024-01-15

This is a quick review on the probability theory. For the sake of brevity, some arguments presented here are only formal with details omitted. For a proper introduction on probability, consult the reference textbooks listed on the course webpage, courses like MATH3280, MATH4050 and MATH5011, or the Department of Statistics.

1 Basic concepts

A *probability space* $(\Omega, \mathcal{F}, P)$ consists of

- *sample space* Ω, a nonempty set of all possible outcomes
- *event space* $\mathcal{F}$, a σ-algebra on Ω which contains events (subsets of Ω) that we would work on
- *probability measure* $P : \mathcal{F} \to [0, 1]$, an assignment of probability (size) to events in $\mathcal{F}$

A *random variable* is a measurable function $X : \Omega \to A \subseteq \mathbb{R}$ that quantifies the outcomes. Shorthand notations like $\{ X \in B \} = \{ \omega \in \Omega \mid X(\omega) \in B \}$ and $\{ X = a \} = \{ \omega \in \Omega \mid X(\omega) = a \}$ are commonly used.

1.1 Conditional and independence

Consider two events $A, B \subseteq \Omega$. If $P(B) \neq 0$, the conditional probability of A given B is $P(A|B) = \frac{P(A\cap B)}{P(B)}$. A, B are independent if $P(A \cap B) = P(A)P(B)$, or $P(A|B) = P(A)$ assuming $P(B) \neq 0$.

If X, Y are two random variables on the same probability space, then X, Y are independent if for all (Borel) sets $A, B \subseteq \mathbb{R}$ the events $\{ X \in A \}$ and $\{ Y \in B \}$ are independent, that is $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$.

Theorem 1.1 (law of total probability)**.** *If $B_1, B_2, \ldots$ are disjoint and $\Omega = \bigcup B_i$, then $P(A) = \sum P(A \cap B_i)$*

2 Distribution functions

For a random variable X, its *cumulative distribution function (CDF)* is $F_X(x) = P(X \leq x)$.

If the range of X is countable, X is a *discrete* random variable, and its *probability mass function (PMF)* is $p_X(x) = P(X = x)$. Commonly the range of a discrete random variable is a subset of integers.

Exercise 2.1. Show that for a discrete random variable X and its range $A = X(\Omega)$, $0 \leq p_X(x) \leq 1$ for all $x \in A$, and $\sum_{x\in A} p_X(x) = 1$

If there exists a function f_X such that $P(a \leq X \leq b) = \int_a^b f_X(x)\,\mathrm{d}x$ for all $a, b \in \mathbb{R}$ with $a \leq b$, then X is a *continuous* random variable, and f_X is its *probability density function (PDF)*. Under some simple conditions, $\frac{\mathrm{d}}{\mathrm{d}x} F_X = f_X$, and $P(X \in A) = \int_A f_X$.

Exercise 2.2. Show that for a continuous random variable X with continuous pdf f_X, $f_X \geq 0$ on $\mathbb{R}$, and $\int_{-\infty}^{\infty} f_X(x)\,\mathrm{d}x = 1$

In this course, we will (usually) assume a random variable is either discrete or continuous.

2.1 Examples

Some classic examples of discrete distributions are

- *binomial distribution* $X \sim \mathrm{B}(n,p)$ with $n \in \mathbb{N}$, $p \in [0,1]$ has pmf $p_X(k) = \binom{n}{k} p^k (1-p)^p$ for $k \in \{ 0, \dots, n \}$. On $n = 1$, $B(1,p)$ is also called *Bernoulli distribution*
- *Poisson distribution* $X \sim \mathrm{Poisson}(\lambda)$ with $\lambda > 0$ has pmf $p_X(k) = \frac{\lambda^k}{k!} e^{-\lambda}$ for $k \in \mathbb{N}$
- *geometric distribution* $X \sim \mathrm{Geom}(p)$ with $p \in [0,1]$ has pmf $p_X(k) = (1-p)^{k-1} p$ for $k \in \mathbb{Z}^+$

Some classic examples of continuous distributions are

- *(continuous) uniform distribution* $X \sim \mathrm{U}(a,b)$ with $-\infty < a < b < \infty$ has pdf $f_X(x) = \frac{1}{b-a} \chi_{[a,b]}(x)$ for $x \in \mathbb{R}$ where $\chi_S(x) = \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{otherwise} \end{cases}$ is the indicator function
- *exponential distribution* $X \sim \mathrm{Exp}(\lambda)$ with $\lambda > 0$ has pdf $f_X(x) = \lambda e^{-\lambda x} \chi_{x \geq 0}$ for $x \in \mathbb{R}$
- *normal distribution / Gaussian distribution* $X \sim \mathrm{N}(\mu, \sigma^2)$ with $\mu \in \mathbb{R}$, $\sigma > 0$ has pdf $f_X(x) = (2\pi\sigma^2)^{-1/2} e^{-(\frac{x-\mu}{\sigma})^2/2}$ for $x \in \mathbb{R}$

2.2 Functions of random variables

If $X_1, \dots, X_n$ are random variables (on the same probability space) and $g : \mathbb{R}^n \to \mathbb{R}$ is a (Borel measurable) function, then $g(X_1, \dots, X_n)$ is also a random variable. To work on such random variable, it is necessary to consider the *joint distribution* with cdf $F_{X_1,\dots,X_n}(x_1, \dots, x_n) = P\left(X_1 \leq x_1, \dots, X_n \leq X_n\right)$.

If $X_1, \dots, X_n$ are independent, the cdf factorizes $F_{X_1,\dots,X_n}(x_1, \dots, x_n) = F_{X_1}(x_1) \dots F_{X_n}(x_n)$. If $X_1, \dots, X_n$ are all discrete, the joint pmf is just the product of all pmf $p_{X_1,\dots,X_n}(x_1, \dots, x_n) = p_{X_1}(x_1) \dots p_{X_n}(x_n)$. Similarly, when all random variables are continuous, the joint pdf is $f_{X_1,\dots,X_n}(x_1, \dots, x_n) = f_{X_1}(x_1) \dots f_{X_n}(x_n)$.

2.2.1 Sum of discrete random variables

Let X, Y be two independent *discrete* random variables with integer values. Then the pmf of $X + Y$ is

$$p_{X+Y}(z) = P(X + Y = z) = \sum_{k \in \mathbb{Z}} P(X = k,\, Y = z - k) = \sum_{k \in \mathbb{Z}} P(X = k) P(Y = z - k) = \sum_{k \in \mathbb{Z}} p_X(k) p_Y(z - k)$$

so $p_{X+Y} = p_X * p_Y$ is the (discrete) convolution of p_X and p_Y.

Example 2.1. Consider independent random variables $X \sim \mathrm{Poisson}(\lambda)$, $Y \sim \mathrm{Poisson}(\mu)$. Then

$$\begin{aligned} p_{X+Y}(z) &= \sum_{k \in \mathbb{Z}} p_X(k) p_Y(z-k) = \sum_{k \in \mathbb{Z}} \chi_{k \geq 0} \frac{\lambda^k}{k!} e^{-\lambda} \chi_{z-k \geq 0} \frac{\mu^{z-k}}{(z-k)!} e^{-\mu} \\ &= e^{-(\lambda+\mu)} \chi_{z \geq 0} \frac{1}{z!} \sum_{k=0}^{z} \frac{z!}{k!(z-k)!} \lambda^k \mu^{z-k} = e^{-(\lambda+\mu)} \frac{(\lambda+\mu)^z}{z!} \chi_{z \geq 0} \end{aligned}$$

so $X + Y \sim \mathrm{Poisson}(\lambda + \mu)$.

By induction, if $X_i \sim \mathrm{Poisson}(\lambda_i)$, $i \in \{ 1, \dots, n \}$ are independent, then $\sum X_i \sim \mathrm{Poisson}(\sum \lambda_i)$

Exercise 2.3. If $X_1, \dots, X_n$ are independent and identically distributed (iid) Bernoulli random variables with common parameter $p \in [0,1]$, show that $X_1 + \dots + X_n \sim \mathrm{B}(n,p)$

2.2.2 Sum of continuous random variables

Let X, Y be two independent *continuous* random variables. Then the cdf of $X+Y$ is

$$F_{X+Y}(z) = P(X+Y \le z) = \int_{-\infty}^{\infty} P(Y \le z-x)\, f_X(x)\ \mathrm{d}x = \int_{-\infty}^{\infty}\int_{-\infty}^{z-x} f_Y(y) f_X(x)\ \mathrm{d}y\ \mathrm{d}x$$

$$\text{so}\quad f_{X+Y}(z) = \frac{\mathrm{d}}{\mathrm{d}z} F_{X+Y}(z) = \int_{-\infty}^{\infty} f_Y(z-x) f(x)\ \mathrm{d}x$$

hence $f_{X+Y} = f_X * f_Y$ is the (continuous) convolution of f_X and f_Y.

Exercise 2.4. Show that the sum $X+Y$ of independent random variables $X \sim \mathrm{N}(\mu_X, \sigma_X^2)$, $Y \sim \mathrm{N}(\mu_Y, \sigma_Y^2)$ is a normal random variable $X+Y \sim \mathrm{N}(\mu_X+\mu_Y, \sigma_X^2+\sigma_Y^2)$

2.2.3 Other combinations

Example 2.2. Suppose $X_i \sim \mathrm{Exp}(\lambda_i)$ for $i \in \{1, \dots, n\}$ with $\lambda_1, \dots, \lambda_n > 0$ are independent random variables. Then on $Y = \min(X_1, \dots, X_n)$, its cdf is

$$\begin{aligned}
F_Y(y) &= P(Y \le y) = 1 - P(Y > y) \\
&= 1 - P(X_1 > y,\ \dots,\ X_n > y) \\
&= 1 - P(X_1 > y) \dots P(X_n > y) \\
&= 1 - (1 - P(X_1 \le y)) \dots (1 - P(X_n \le y)) \\
&= 1 \cdot \chi_{y \ge 0} - e^{-\lambda_1 y} \dots e^{-\lambda_n y} \chi_{y\ge 0} = (1 - e^{-y\sum \lambda_i})\chi_{y \ge 0}
\end{aligned}$$

because $P(X_i \le y) = \chi_{y\ge 0}(1 - e^{-\lambda_i y})$. So $Y \sim \mathrm{Exp}(\sum \lambda_i)$.

Note that this also implies $P(X_1 = \min(X_1, \dots, X_n)) = \lambda_1 / \sum \lambda_i$ since on $Z = \min(X_2, \dots, X_n) \sim \mathrm{Exp}(\lambda)$ with $\lambda = \lambda_2 + \dots + \lambda_n$, X_1, Z are independent and

$$\begin{aligned}
P(X_1 = Y) = P(X_1 \le Z) &= \int_{-\infty}^{\infty} P(X_1 \le z)\, f_Z(z)\ \mathrm{d}z \\
&= \int_0^{\infty} (1 - e^{-\lambda_1 z})\lambda e^{-\lambda z}\ \mathrm{d}z = 1 - \frac{\lambda}{\lambda_1 + \lambda} = \frac{\lambda_1}{\lambda_1+\lambda} = \frac{\lambda_1}{\sum \lambda_i}
\end{aligned}$$

Exercise 2.5. Suppose $X_1 \sim \mathrm{Exp}(\lambda_1)$, $X_2 \sim \mathrm{Exp}(\lambda_2)$ with $\lambda_1, \lambda_2 > 0$ are independent random variables. Find the pdf of $\max(X_1, X_2)$.

For random variables that are not independent with each other, it is often necessary to consider the full joint distribution.

2.3 Moments

The *expected value* of a random variable X is $\mathrm{E}(X) = \sum x p_X(x)$ if X is discrete, and $\mathrm{E}(X) = \int_{-\infty}^{\infty} x f_X(x)\ \mathrm{d}x$ if X is continuous, assuming the value is well-defined.

Assuming $\mu = \mathrm{E}(X)$ is well-defined, The *variance* of a random variable X is $\mathrm{Var}(X) = \mathrm{E}\left((X-\mu)^2\right)$, if it is also well-defined.

Note that these two moments may not be well-defined, e.g. the Cauchy distribution $X \sim \mathrm{Cauchy}(\mu, \gamma)$ with $\mu \in \mathbb{R}$, $\gamma > 0$ which has pdf $f_X(x) = \frac{\gamma}{\pi}((x-\mu)^2 + \gamma^2)^{-1}$.

Exercise 2.6. Show that the Cauchy distribution does not have a well-defined expected value

Properties of expected values: if X, Y are random variables, then assuming all quantities are well-defined,

- $\mathrm{E}(cX) = c\mathrm{E}(X)$ for all $c \in \mathbb{R}$
- $\mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y)$
- $\mathrm{Var}(X) = \mathrm{E}\left(X^2\right) - \mathrm{E}(X)^2 \ge 0$

- $\mathrm{Var}\left(cX\right) = c^2\mathrm{Var}\left(X\right)$ for all $c \in \mathbb{R}$
- (law of total expectation) if $A_1, A_2, \ldots$ are disjoint and $\bigcup A_i = \Omega$, then $\mathrm{E}\left(X\right) = \sum \mathrm{E}\left(X|A_i\right) P\left(A_i\right)$

Example 2.3. Let X, Y are independent discrete random variables of integer value with finite expected values and variances, then

$$\mathrm{E}\left(XY\right) = \sum \mathrm{E}\left(XY|Y=k\right) P\left(Y=k\right) = \sum k\mathrm{E}\left(X|Y=k\right) P\left(Y=k\right) = \sum k\mathrm{E}\left(X\right) P\left(Y=k\right) = \mathrm{E}\left(X\right)\mathrm{E}\left(Y\right)$$

Also,

$$\begin{aligned}
\mathrm{Var}\left(X+Y\right) &= \mathrm{E}\left((X+Y)^2\right) - \mathrm{E}\left(X+Y\right)^2 \\
&= (\mathrm{E}\left(X^2\right) + 2\mathrm{E}\left(XY\right) + \mathrm{E}\left(Y^2\right)) - (\mathrm{E}\left(X\right)^2 + 2\mathrm{E}\left(X\right)\mathrm{E}\left(Y\right) + \mathrm{E}\left(Y\right)^2) \\
&= (\mathrm{E}\left(X^2\right) - \mathrm{E}\left(X\right)^2) + 2(\mathrm{E}\left(XY\right) - \mathrm{E}\left(X\right)\mathrm{E}\left(Y\right)) + (\mathrm{E}\left(Y^2\right) - \mathrm{E}\left(Y\right)^2) \\
&= \mathrm{Var}\left(X\right) + \mathrm{Var}\left(Y\right)
\end{aligned}$$

These also hold for continuous random variables.

Exercise 2.7. Verify that

	expected value	variance
$\mathrm{B}(n,p)$	np	$np(1-p)$
$\mathrm{Poisson}(\lambda)$	λ	λ
$\mathrm{Geom}(p)$	$1/p$	$(1-p)/p^2$
$\mathrm{U}(a,b)$	$(a+b)/2$	$(b-a)^2/12$
$\mathrm{Exp}(\lambda)$	$1/\lambda$	$1/\lambda^2$
$\mathrm{N}(\mu,\sigma^2)$	μ	σ^2