Lecture 0: Preliminary

0.1 STAT4351 Review

- sample space, probability of events, Kolmogorov's axioms, independence and dependence, Bayesian methods
- discrete vs. continuous random variables
- PMF and PDF of univariate and multivariate random variables
- expectation, variance, mgf, covariance, etc.
- special probability distributions
- functions of random variables, transformations for one and several variables
- law of large numbers, central limit theorem, sampling distributions

0.2 Mathematical Foundations

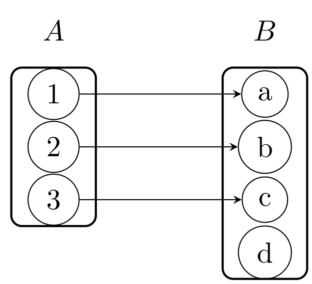
Definition 0.1 A function is **injective** (one-to-one) if different elements in A are mapped to different elements in B. Formally,

$$\forall x_1, x_2 \in A, \quad f(x_1) = f(x_2) \implies x_1 = x_2 \quad [\text{ or }] \quad \forall x_1, x_2 \in A, \quad x_1 \neq x_2 \implies f(x_1) \neq f(x_2)$$

Definition 0.2 A function is **surjective** (onto) if every element in B has at least one preimage in A. Formally,

$$\forall y \in B, \quad \exists x \in A \text{ such that } f(x) = y$$

Example 0.1 A function $f : A \to B$ that is injective, but not surjective:



Example 0.2 The function $g : \mathbb{R} \to \mathbb{R}$, $g(x) = x^2$ is not injective and not surjective.

Proof. Counter example for injective: $g(1) = g(-1) = 1$. Counter example for surjective: $g(x) \geq 0$ for all $x \in \mathbb{R}$. $\square$

Definition 0.3 A function is **bijective** (one-to-one correspondence) if it is both injective and surjective. Formally,

$$\forall y \in B, \quad \exists! x \in A \text{ such that } f(x) = y \qquad (\text{! means “unique”})$$

Alternatively, $f : A \to B$ is **bijective** if and only if $\exists g : B \to A$ (inverse of f) such that $g(f(x)) = x$, $\forall x \in A$, and $f(g(y)) = y$, $\forall y \in B$.

Example 0.3 Consider the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = 2x + 3$.

(a) Injective?

Proof. Suppose $f(x_1) = f(x_2)$, then $2x_1 + 3 = 2x_2 + 3 \implies x_1 = x_2$ so f is injective. □

(b) Surjective?

Proof. For any $y \in \mathbb{R}$, solving $y = 2x + 3$ for x gives $x = \frac{y-3}{2}$. Since a solution for x exists $\forall y \in \mathbb{R}$, f is surjective. □

(c) Bijective?

Proof. Since f is both injective and surjective, it is bijective. □

0.3 Moment Generating Function (mgf)

As the name suggests, mgfs are used to find the “moments” of a distribution.

Definition 0.4 If $\mathbb{E}(|X^k|) < +\infty$, we define the following quantity

$$\mathbb{E}(X^k) = \begin{cases} \sum_x x^k f(x), & X \text{ is discrete random variable,} \\ \int_{-\infty}^{\infty} x^k f(x)\, dx, & X \text{ is continuous random variable.} \end{cases}$$

as the ***k*th moment of a random variable *X*** $(k = 1, 2, \ldots)$, denoted $m_k(X)$ or m_k.

Definition 0.5 The **moment generating function** of X is defined as

$$M_X(t) = \text{mgf}_X(t) = \mathbb{E}(e^{tX})$$

if the expectation exists for $-h < t < h$, for some constant $h > 0$.

Remark 0.1 The Taylor Series expansion of e^{tX} is given by

$$e^{tX} = 1 + tX + \frac{(tX)^2}{2!} + \cdots + \frac{(tX)^k}{k!} + \cdots$$

Then, taking the expectation we know

$$\begin{aligned}\mathbb{E}(e^{tX}) = M_X(t) &= 1 + t\mathbb{E}(X) + \frac{t^2}{2!}\mathbb{E}(X^2) + \cdots + \frac{t^k}{k!}\mathbb{E}(X^k) + \cdots \\ &= 1 + tm_1 + \frac{t^2}{2!}m_2 + \cdots + \frac{t^k}{k!}m_k + \cdots\end{aligned}$$

where m_k denotes the k^{th} moment, and can be obtained by differentiating the mgf k-times and setting $t = 0$.

Remark 0.2 Properties of mgfs:

1. (Distribution determination) If two random variables X, Y has the same mgf for all $-h < t < h$, for some constant $h > 0$, then X and Y follow the same distribution (i.e., $X \overset{d}{\equiv} Y$).
2. (Independence) Two random variables X, Y are independent if and only if

$$\text{mgf}_{X,Y}(t, s) = \mathbb{E}(e^{tX+sY}) = \text{mgf}_X(t)\text{mgf}_Y(s).$$

0.4 Cumulant generating function (CGF)

If the moment generating function $M(t)$ of the random variable X exists, taking the logarithm of the mgf is known as the cumulant generating function (CGF) of the random variable X.

$$C(t) = \log M(t) = \log E(e^{tX}), \quad -h < t < h\,(\exists h > 0)$$

It is often denoted by cgf. The Taylor expansion of the cumulant generating function is

$$C(t) = \sum_{r=0}^{\infty} \frac{C^{(r)}(0)}{r!}t^r, \quad -h < t < h\,(\exists h > 0), \quad r = 1, 2, \ldots$$

where $C(0) = 0$, and the coefficients $C^{(r)}(0)$ are called the r-th cumulants of X, denoted by $c_r(X)$ or simply c_r. That is,

$$C(t) = \sum_{r=1}^{\infty} \frac{c_r(X)}{r!}t^r, \quad -h < t < h\,(\exists h > 0).$$

The relationship between the cumulants c_r and the moments m_k can be derived from

$$\log\left(1+\sum_{k=1}^{\infty}\frac{m_k}{k!}t^k\right)=\sum_{r=1}^{\infty}\frac{c_r}{r!}t^r,\quad -h<t<h\,(\exists h>0)$$

Using the expansion for $\log(1+A)$ as follows

$$\log(1+A)=A-\frac{A^2}{2}+\frac{A^3}{3}-\cdots,\quad (-1<A<1),$$

where $A=m_1t+m_2t^2/2!+m_3t^3/3!+\cdots$, and expanding in powers of t, we get:

$$c_1=m_1,\quad c_2=2!\left(m_2/2!-(m_1)^2/2!\right)=m_2-(m_1)^2,\ldots$$

Thus, the cumulants c_1 and c_2 represent the mean and variance, respectively.

There are some interesting properties for the standardized random variable $Z=\frac{X-\mu}{\sigma}$.

- The relationship between the moments $m_r(Z)$ and cumulants $c_r(Z)$ is given by

$$\begin{aligned} c_3(Z)&=m_3(Z)=\mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]\\ c_4(Z)&=m_4(Z)-3=\mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^4\right]-3, \end{aligned}$$

  where $c_3(Z)$ is the third cumulant, and $c_4(Z)$ is the fourth cumulant, which define the skewness and kurtosis of X, respectively.

- The relationship between the cumulant generating functions of Z and X, and between the cumulants of Z and X, is given by:

$$c_r\left(\frac{X-\mu}{\sigma}\right)=\frac{c_r(X)}{\sigma^r},\quad r=3,4,\ldots.$$

Proof. [Sketch of Proof] The skewness and kurtosis are defined through the standardization of X.

$$M_Z(t):=1+\sum_{k=1}^{\infty}\frac{m_kt^k}{k!},\quad A:=\sum_{k=1}^{\infty}\frac{m_kt^k}{k!}$$

Then:

$$\mathbb{E}[Z]=m_1=0,\quad \mathrm{Var}(Z)=1,\quad m_2-m_1^2=m_2,$$

and therefore,

$$A=\frac{1}{2}t^2+\frac{m_3}{3!}t^3+\frac{m_4}{4!}t^4+\cdots.$$

Now, the cumulant generating function (CGF) of Z, denoted by:

$$\mathrm{cgf}_Z(t) := \sum_{r=1}^{\infty} \frac{c_r t^r}{r!}$$

can be derived by comparing the coefficients as follows

$$\mathrm{cgf}_Z(t) = \log M_Z(t) = \log(1 + A) = A - \frac{A^2}{2} + \frac{A^3}{3} - \cdots .$$

and we can conclude that

$$c_3 = 3! \cdot \frac{m_3}{3!} = m_3, \quad c_4 = 4! \left(\frac{m_4}{4!} - \frac{1}{2} \cdot \frac{1}{(2!)^2} \right) = m_4 - 3.$$

For the second bullet point, from the definition of the mgf, we have

$$M_Z(t) = E\left[e^{t\frac{X-\mu}{\sigma}}\right] = E\left[e^{-\frac{\mu t}{\sigma}} e^{\frac{t}{\sigma}X}\right] = e^{-\frac{\mu t}{\sigma}} M_X\left(\frac{t}{\sigma}\right)$$

Now, taking the logarithm of both sides

$$\mathrm{cgf}_Z(t) = \log M_Z(t) = -\frac{\mu t}{\sigma} + \log M_X\left(\frac{t}{\sigma}\right) = -\frac{\mu t}{\sigma} + \mathrm{cgf}_X\left(\frac{t}{\sigma}\right)$$

From this, we get

$$c_r(Z) = \mathrm{cgf}_Z^{(r)}(0) = \frac{1}{\sigma^r} \cdot \mathrm{cgf}_X^{(r)}(0) = \frac{1}{\sigma^r} \cdot c_r(X), \quad r = 3, 4, \ldots$$

□