BE/Bi 103B Recitation 2

Sophie Walton

Review of Basic Concepts / Definitions

① Sample Space: Set of all possible outcomes of an "experiment". denoted as S

② Distribution: mapping of outcomes in sample space → probabilities $x_i \in S \rightarrow p_i$

ⓘ all probabilities p_i are $\in [0,1]$

② $\sum p_i = 1$.

③ Random: a function mapping outcomes in sample space to unique real #s.

④ Independence: Two events $A, B \in S$ are independent (denoted $A \perp\!\!\!\perp B$) if

$$\mathbb{P}(AB) = P(A) \cdot \mathbb{P}(B)$$

⑤ Conditional Probability: If $P(B) > 0$ then

$$\underbrace{\mathbb{P}(A \mid B)}_{\text{probability of A occuring given that B occured}} = \frac{\mathbb{P}(A, B)}{\mathbb{P}(B)}$$

← probability they both occur

← probability B occurs

Note: If $A \perp\!\!\!\perp B$ $\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A, B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A)\cancel{\mathbb{P}(B)}}{\cancel{\mathbb{P}(B)}} = \mathbb{P}(A)$:)

⑥ Baye's Formula:

ⓐ $\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A, B)}{\mathbb{P}(B)}$

ⓑ $\mathbb{P}(B \mid A) = \frac{\mathbb{P}(B, A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A, B)}{\mathbb{P}(A)} \Rightarrow$ ⓒ $\mathbb{P}(A, B) = \mathbb{P}(B \mid A)\,\mathbb{P}(A)$

Thus we can plug ⓒ into ⓐ and obtain

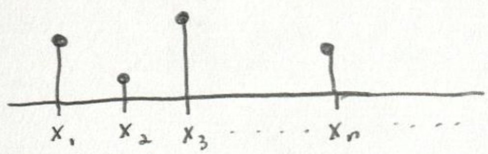
$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A)\,\mathbb{P}(A)}{\mathbb{P}(B)}$$:)

Discrete Random Variables → implies that possible of outcomes are countable! or finite!

The Probability mass function (PMF): describes probabilities of outcomes for discrete random variables. For a general discrete random variable X, the probability mass function is

$$p(x_i) = \mathbb{P}(X = x_i) \quad \text{where } x_i \in \underbrace{\{x_1, \dots\}}_{\text{sample space for } X}$$

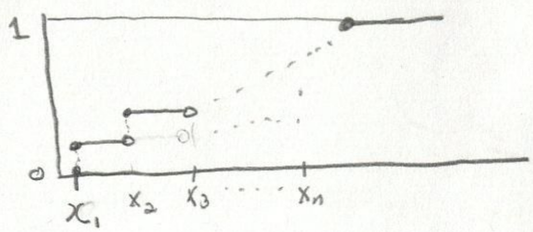
And $\sum_{i=1}^{\infty} p(x_i) = 1$



← example plot of pmf

The cummolative distribution function is an equivalent manner of specifying probabilities. It is defined as

$$F_X(x) = \mathbb{P}(X \leq x) = \sum_{x_i \leq x} p(x_i)$$



Continuous Random Variables → implies that set of possible outcomes are infinite

The probability density function is the analog of the PMF except it is for continuous random variables.

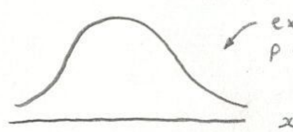
Since our random variable is continuous, and thus can take on ∞ many values we have $\mathbb{P}(X = x) = 0$ where $x \in$ sample space of $X = S$

Instead we consider

$$\mathbb{P}(X \in A) = \int_A p(t)\,dt \quad \text{where } A \subset S.$$

$p(x)$ is the PDF and $\int_S p(x)\,dx \doteq \int_{-\infty}^{\infty} p(x)\,dx = 1.$

Unlike the PMF, the PDF is smooth.



← example PDF

The CDF for a continuous random variable is defined in the same way as for discrete random variables. For

$$F_x(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} p(t)\,dt$$

→ Note: Evaluating the PDF at a point is not the probability of that value.

You can view it as $\mathbb{P}\left(x - \frac{\varepsilon}{2} \leq X \leq x + \frac{\varepsilon}{2}\right) = \int_{x-\frac{\varepsilon}{2}}^{x+\frac{\varepsilon}{2}} p(x)\,dx \approx \varepsilon\, p(x)$. ε is small. It is like the probability that you will be "near" x.

Expectation Values

In general an expectation value is denoted

$$\mathbb{E}_x[g(x)]$$; where $g(x)$ is a function applied to the random variable.

In discrete world this is

$$\mathbb{E}_x[g(x)] = \sum_{x_i \in S} g(x_i) \underbrace{p(x_i)}_{\text{PMF for } X}$$

In continuous world this is

$$\mathbb{E}_x[g(x)] = \int_{-\infty}^{\infty} g(t)p(t)dt$$

The n^{th} moment is $\mathbb{E}_x[x^n]$

The mean is the 1^{st} moment or

$$\mu = \mathbb{E}_x[x] = \sum_{x_i \in S} x_i p(x_i) \leftarrow \text{if discrete}$$

$$\mu = \mathbb{E}_x[x] = \int_{-\infty}^{\infty} t\,p(t)\,dt \leftarrow \text{if continuous}$$

The variance is

$$\mathbb{E}_x[(x - \overset{\text{the mean}}{\mu})^2] = \mathbb{E}_x[x^2] - (\mathbb{E}_x[x])^2$$

and the standard deviation is

$$\sigma = \sqrt{\mathbb{V}[X]}$$

Discrete Random Variables

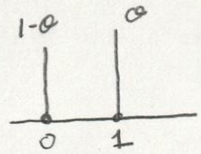
① **Bernoulli:** Describes the outcome of a single trial with probability θ of success, and $1-\theta$ of failure.

(Like A single coin flip)

For $X \sim \text{Bernoulli}(\theta)$ we have

$$p(X=x) = \begin{cases} 1-\theta, & x=0 \text{ (failure)} \\ \theta & x=1 \text{ (success)} \end{cases}$$

PMF



Mean

$\mathbb{E}_X[x] = \theta$.

Variance $\mathbb{V}_X[x] = \theta(1-\theta)$

② **Binomial:** We perform N independent identically distributed (iid) Bernoulli trials with probability of success θ.

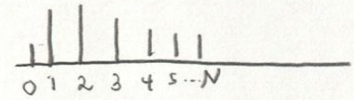
X = the number of successes. $X \sim \text{Bin}(N, \theta)$

(For example the number of heads you get in N coin flips)

$$p(X=x) = \binom{N}{x}\theta^x(1-\theta)^{N-x}$$

$$\binom{N}{x} = \frac{N!}{x!(N-x)!}$$

PMF



Mean

$\mathbb{E}_X[x] = N\theta$

$\mathbb{V}_X[x] = N\theta(1-\theta)$

③ **Geometric:** We perform a series of iid Bernoulli trials. The number of "failures" until we get a "success" is geometrically distributed.

The number of coin flips until I get a head is geometrically distributed

For $X \sim \text{Geometric}(\theta)$ ← probability of success for each trial

$$p(X=x) = \underbrace{(1-\theta)^x}_{x \text{ failures}}\ \theta \leftarrow 1 \text{ success}$$

Mean:

$\mathbb{E}_X[x] = \frac{1-\theta}{\theta}$

Variance

$\mathbb{V}_X[x] = \frac{1-\theta}{\theta^2}$

④ **Poisson:** Consider a binomially distributed random variable with $N \to \infty$ and $\theta \to 0$ such that $N \cdot \theta = \lambda = \text{const}$.

The Poisson is often used to model "rare" events

$$\mathbb{P}(X=x) = \binom{N}{x}\theta^x(1-\theta)^{n-x} = \frac{N!}{x!(N-x)!}\left(\frac{\lambda}{N}\right)^x\left(1-\frac{\lambda}{n}\right)^{N-x}$$

$$= \underbrace{\frac{N(N-1)\cdots(N-x+1)}{N^n}}_{\to 1 \text{ as } N\to\infty}\ \frac{\lambda^x}{x!}\ \frac{\overbrace{\left(1-\frac{\lambda}{N}\right)^n}^{\to e^{-\lambda}}}{\underbrace{\left(1-\frac{\lambda}{N}\right)^x}_{\to 1}} \to \boxed{e^{-\lambda}\frac{\lambda^x}{x!}}$$

Mean: λ

Variance: λ

The Poisson Process

$N_t = \#$ of "events" by time t

$\{N_t, t \geq 0\}$

The poisson is a type of counting process with rate λ s.t. $\lambda > 0$

It satisfies the following:

① $N_0 = 0$

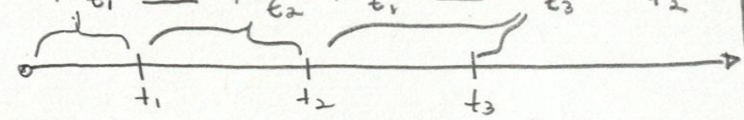
② $\{N_t\}$ has stationary + independent increments

(i) Stationary means that the distribution of the number of events in a time interval $(s, s+t]$ only depends on t, the length of the interval

(ii) independence of increments means that if two intervals are disjoint, then the distribution of the number of events in those intervals are independent from each other.

For example, let $t_1 < t_2 < t_3$. Then

$$N_{t_1} \perp\!\!\!\perp N_{t_2} - N_{t_1} \perp\!\!\!\perp N_{t_3} - N_{t_2}$$



Reminder

$\perp\!\!\!\perp$ means independent

is true:

③ For all $s, t \geq 0$ $\quad N_{t+s} - N_s \sim \text{Poisson}(\lambda t)$

$$\left(\mathbb{P}(N_{t+s} - N_s = n) = \exp(-\lambda t)\,\frac{(\lambda t)^n}{n!}\right)$$

Thus $\mathbb{E}[N_t] = \lambda t$

Interarrival times

Let T_1 be the time of the 1st event.

Let T_n be the time between event $n-1$ and event n.

Then the arrival times are $\{T_n, n = 1, 2, \ldots\}$

What is the distribution of T_n?

- $n=1$ $\quad \mathbb{P}(T_1 > t) = \mathbb{P}(N_t = 0) = \lambda e^{-\lambda t}$

- In general it can be shown that

$$\mathbb{P}(T_n > t) = \lambda e^{-\lambda t}$$

This leads us to the definition of an Exponential Random variable. An exponential Random variable is characterized by a rate parameter λ.

The PDF is $\quad p(x) = \lambda e^{-\lambda x}$

It is MEMORYLESS. The amount of time we will continue to wait for an arrival is independent of how long we have been waiting.

So In general, the exponential can be used to model waiting times between rare events.

Now Let's talk about arrival times.

Let S_n denote the time we wait for arrival n.
For the n^{th} arrival to occur we must have arrivals $n-1, n-2, \ldots, 1$ all occur. Remember that $T_1, T_2, \ldots, T_n$ are the arrival waiting times between these events, thus

$$S_n = T_1 + T_2 + \ldots + T_n.$$

Remember that $T_1 \ldots T_n \sim Exp(\lambda)$

It can then be shown that the PDF of S_n is

$$p(t; n, \lambda) = \frac{1}{\Gamma(n)} \frac{(\lambda t)^n}{t} \exp(-\lambda t)$$

Thus S_n is a Gamma distributed random variable.

It can be used to model the time for any multistep process where each step happens at the same rate.

Important Moments For Exponential + Gamma Distributions

Exponential:

$$\text{mean} = \frac{1}{\lambda}$$

$$\text{variance} = \frac{1}{\lambda^2}$$

Gamma:

$$\text{mean} = \frac{n}{\lambda}$$

$$\text{variance} = \frac{n}{\lambda^2}$$

(Note: n can be non integer!!)

The Normal Distribution

This is probably the most famous continuous distribution. The story for the normal distribution is that many quantities that are controlled by many subprocesses are often approximatly normally distributed as long as none of the subprocesses has a large variance. It is parameterized by μ (the mean) and σ (the standard deviation). It is defined for all real numbers and its PDF is:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-(x-\mu)^2/2\sigma^2\right) \text{ for } X \sim N(\mu, \sigma)$$

The Central Limit Theorem

Let $X_1, X_2, \ldots X_n$ be a sequence of iid (independent identically distributed) random variables with mean $\mu = \mathbb{E}[X_i]$ and variance $\sigma^2 = \mathbb{V}[X_i]$

Classic

Then $\bar{X} \overset{\cdot}{\sim} N(\mu, \frac{\sigma}{\sqrt{n}})$

$\overset{\cdot}{\sim}$ approximatly distributed as

General

There are other more general variants of the CLT. In the end it is because of the CLT that so many quantities that are the sum of many subprocesses with low variance tend to be normally distributed.

Multivariate Normal

Let $X_1, X_2, \ldots, X_n$ be normally distributed random variables (vectorial random variables for entries)

Then $\underset{\sim}{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$ is a Gaussian random vector, $\underset{\sim}{X} \sim (\underset{\sim}{\mu}, \underset{\approx}{\Sigma})$

$$\underset{\sim}{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} \qquad \underset{\approx}{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & & \\ \vdots & & \ddots & \\ \sigma_{n1} & & & \sigma_n^2 \end{pmatrix}$$

covariance $= \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)]$

$$\underset{\approx}{\Sigma}_{ij} = \begin{cases} \sigma_{ij} & i \neq j \\ \sigma_i^2 & i = j \end{cases}$$

if $X_i \perp\!\!\!\perp X_j$ then $\sigma_{ij} = 0$

The PDF is

$$p(\underset{\sim}{X}) = \frac{1}{\sqrt{(2\pi)^n \det \underset{\approx}{\Sigma}}} \exp\left[-\frac{1}{2}(\underset{\sim}{X} - \underset{\sim}{\mu})^T \underset{\approx}{\Sigma}^{-1} (\underset{\sim}{X} - \underset{\sim}{\mu})\right]$$