

Brief Theory of Probability: Notes from MATH 431

Compiled by Harry Luo

Contents

1 Random Variables	2
1.1 Discrete random variable	2
1.1.1 Probability Mass Function (pmf)	2
1.2 continuous Random Variables	3
1.2.1 Probability Density Function (pdf)	3
1.2.2 Cumulative Distribution Function (cdf)	3
1.3 Expectation and Variance	4
1.3.1 Expectation	4
1.3.2 Expectation of a function of a random variable	5
1.3.3 Moments, and moment generating function	5
1.3.4 Variance	5
2 continuous Distribution	6
2.1 Uniform Distribution	6
2.2 Normal (Gaussian) Distribution	6
2.2.1 standard normal distribution	6
2.2.2 normal distribution (generalized)	6
3 Approximations of Binomial Distribution	7
3.1 Central limit theorem (approximation with normal distribution)	7
3.1.1 continuity correction	7
3.1.2 Law of large numbers	7
3.1.3 Confidence interval	7
3.2 Poisson Distribution	8
3.2.1 Poisson r.v.	8
3.2.2 Law of rare events	8
3.3 Exponential Distribution	8
4 Joint Distribution	9
4.1 discrete joint distribution	9
4.2 Continuous joint distribution	9
4.3 Independent joint random variables	9

1 Random Variables

Properties of Random Variables	
Discrete	Continuous
Probability mass function $p_X(k) = P(X = k)$	Probability density function $f_X(x)$
$P(X \in B) = \sum_{k: k \in B} p_X(k)$	$P(X \in B) = \int_B f_X(x) dx$
Cumulative distribution function $F_X(a) = P(X \leq a)$	
$F_X(a) = \sum_{k: k \leq a} p_X(k)$ F_X is a step function.	$F_X(a) = \int_{-\infty}^a f(x) dx$ F_X is a continuous function.
$P(X < a) = \lim_{t \rightarrow a^-} F(t) = F(a-)$ $P(X = a) = F(a) - \lim_{t \rightarrow a^-} F(t) = F(a) - F(a-)$	
$E(X) = \sum_k k p_X(k)$	$E(X) = \int_{-\infty}^{\infty} x f(x) dx$
$E(aX + b) = aE[X] + b$	
$E[g(X)] = \sum_k g(k) p_X(k)$	$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx$
$\text{Var}(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2$	
$\text{Var}(aX + b) = a^2 \text{Var}(X)$	

1.1 Discrete random variable

Discrete random variables are random variables that can take on a countable number of values. It comes naturally from discrete, finite or infinitely countable sample spaces. (As briefly discussed in sec.discreteSampleSpace)

For $A = \{k_1, k_2, \dots\}$ s.t. random variable $X \in A$, or $P(X \in A) = 1$, X is a random variable, with possible values k_1, k_2, \dots and $P(X = k_n) > 0$

1.1.1 Probability Mass Function (pmf)

The PMF is a function that defines the probability distribution for a discrete random variable. It gives the probability of the random variable taking on each possible value. The PMF, denoted as

$$p_X(k) = P(X = k), \text{ where } k \text{ are possible values of } X \quad (1)$$

It is a function of k , and

$$p_X : S \rightarrow [0, 1], \quad (2)$$

where:

S is the support set, i.e., the set of all possible values that the discrete random variable X can take. $[0, 1]$ represents the range of the function, as probabilities are always between 0 and 1. For each value k in the support set S , the PMF assigns a

probability $p_X(k)$, which represents the likelihood of the random variable X taking the value k .

The PMF satisfies the following properties:

Non-negativity: $p_{X(k)} \geq 0$ for all k in S .

Total probability: $\sum_k p_{X(k)} = 1$ where the sum is taken over all k in S .

Example: For a fair six-sided die, the PMF would be $P(X = x) = \frac{1}{6}$ for $x = 1, 2, 3, 4, 5, 6$. Or more elegantly,

$$p_X(k) = \frac{1}{6}, \text{ for every } k \in \{1, 2, 3, 4, 5, 6\} \quad (3)$$

1.2 continuous Random Variables

Not rigorously defined in this class, but a continuous random variable is one that can take on any value in a range. The probability of a continuous random variable taking on a specific value is 0. It came naturally from continuous sample spaces. The probability is assigned to intervals of values, and they are assigned by the **probability density function**.

1.2.1 Probability Density Function (pdf)

continuous r.v are defined in this class by having a probability density function.

A random variable X is continuous if there exists a function $f(x)$ such that

$$\int_{-\infty}^{\infty} f(x) dx = 1, f(x) > 0 \text{ everywhere} \quad (4)$$

$$\text{and } P(X \leq b) = \int_{-\infty}^b f(x) dx \Leftrightarrow P(a \leq X \leq b) = \int_a^b f(x) dx$$

1.2.2 Cumulative Distribution Function (cdf)

cdf of a r.v. is defined as

$$F(x) = P(X \leq x) \quad (5)$$

and it follows that

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a) \quad (6)$$

- Continuous r.v.

it looks suspiciously like an indefinite integral, and when we are dealing with continuous r.v., it is.

$$F(s) = P(X \leq s) = \int_{-\infty}^s f(x) dx$$

Recall the fundamental theorem of calculus,

$$F'(x) = f(x), \quad (7)$$

so the pdf is the derivative of the cdf.

- Discrete r.v.

pmf and cdf is connected by

$$F(x) = P(X \leq s) = \sum_{k \leq x} p_{X(k)} \quad (8)$$

where the sum is taken over all k such that $k \leq x$.

In english, the cdf is the sum of the pmf up to the value x , or “compound probability thus far”

If the cdf graph is stepped (piecewise constant), it is a discrete r.v. If it is continuous except at several points, it is a continuous r.v.

1.3 Expectation and Variance

1.3.1 Expectation

1. Exp of discrete r.v. is defined as

$$E(X) = \sum_k kP(X = k) \quad (9)$$

where the sum is taken over all possible values of X . It is the weighted average of the possible values of X , where the weights are given by the probabilities.

Expectation is a linear operator, i.e.

$$E(aX + b) = aE(X) + b \quad (10)$$

for any constants a and b .

• exp of **Bernoulli** r.v. is

$$E(X) = p \quad (11)$$

where p is the probability of success.

• exp of **binomial** r.v. is

$$E(X) = np \quad (12)$$

where n is the number of trials and p is the probability of success.

• exp of **geometric** r.v. is

$$E(X) = \frac{1}{p} \quad (13)$$

where p is the probability of success.

1. Exp of continuous r.v. is defined as

$$E(X) = \int_{-\infty}^{\infty} xf(x) dx \quad (14)$$

where the integral is taken over the entire range of possible values of X . It is the weighted average of the possible values of X , where the weights are given by the probability density function.

• exp of **uniform** r.v. is

$$E(X) = \frac{a + b}{2} \quad (15)$$

where a and b are the lower and upper bounds of the interval.

1.3.2 Expectation of a function of a random variable

When we have a function of a random variable, we can find the expectation of that function by applying the function to each possible value of the random variable and taking the weighted average of the results.

- if X is a discrete r.v. with pmf $p_X(k)$, and g is a function of X , then

$$E(g(X)) = \sum_k g(k)p_{X(k)} \quad (16)$$

- if X is a continuous r.v. with pdf $f(x)$, and g is a function of X , then

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)f(x) dx \quad (17)$$

1.3.3 Moments, and moment generating function

1. The **nth moment** of the random variable X is the expectation $E(X^n)$.

- X as discrete r.v. with pmf $p_X(k)$, the nth moment is

$$E(X^n) = \sum_k k^n p_{X(k)} \quad (18)$$

- X as continuous r.v. with pdf $f(x)$, the nth moment is

$$E(X^n) = \int_{-\infty}^{\infty} x^n f(x) dx \quad (19)$$

2. The **moment generating function** of a

- discrete random variable X is defined as

$$M_X(t) = E(e^{tX}) = \sum_k e^{tk} p_{X(k)} \quad (20)$$

- continuous random variable X is defined as

$$M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x) dx \quad (21)$$

It is a function of t .

We can easily find the nth moment of X by taking the nth derivative of the moment generating function with respect to t and evaluating it at $t = 0$. i.e.

$$E(X^n) = \frac{d}{dt} M_X(t=0) \quad (22)$$

1.3.4 Variance

The variance of a random variable X is a measure of how much the values of X vary around the mean. It is defined as the expectation of the squared deviation of X from its mean. i.e.

$$\sigma^2 = \text{Var}(X) = E((X - E(X))^2) \quad (23)$$

alternatively,

$$\text{Var}(X) = E(X^2) - (E(X))^2 \quad (24)$$

Variance is not a linear operator, i.e.

$$\text{Var}(aX + b) = a^2 \text{Var}(X) \quad (25)$$

for any constants a and b.

1. variance of bournoli r.v. is

$$p(1 - p) \quad (26)$$

2. variance of binomial r.v. is

$$np(1 - p) \quad (27)$$

3. variance of geometric r.v. is

$$\frac{1 - p}{p^2} \quad (28)$$

4. variance of uniform r.v. is

$$\frac{(b - a)^2}{12} \quad (29)$$

2 continuous Distribution

Based on different pdf, we have different behaviors of random variables. We call them distributions.

2.1 Uniform Distribution

r.v. X has the uniform distribution on the interval [a,b] if its pdf is

$$f(x) = \begin{cases} \frac{1}{b - a} & \text{for } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases} \quad (30)$$

2.2 Normal (Gaussian) Distribution

2.2.1 standard normal distribution

r.v. Z has the Standard normal distribution if its pdf is

$$f(z) = \varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \quad (31)$$

where z is the standard normal r.v. and phi is the standard normal pdf. It's abbreviated as $Z \sim N(0, 1)$ where 0 is the mean and 1 is the variance.

- The **cdf** of the standard normal distribution is denoted as

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^z \varphi(z) dz \quad (32)$$

Check for table for values of $\Phi(z)$

2.2.2 normal distribution (generalized)

two parameters: the mean μ and the variance σ^2 . The pdf of a normal distribution is given by the formula:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(x - \mu)^2}{2\sigma^2} \right] \quad (33)$$

abbreviated as $X \sim N(\mu, \sigma^2)$

- Linearity of normal distribution

If $X \sim N(\mu, \sigma^2)$, $Y = aX + b$, then $Y \sim N(a\mu + b, a^2\sigma^2)$

- **normalization of normal distribution** For $X \sim N(\mu, \sigma^2)$, we can standardize it to $Z \sim N(0, 1)$ by $Z = \frac{X - \mu}{\sigma}$

3 Approximations of Binomial Distribution

Recall: **Binomial distribution** is the distribution of the *number of successes* of n independent Bernoulli trials. It has two parameters: the number of trials n and the probability of success p .

Depending on the probability of success p and the number of trials n , the binomial distribution can be approximated by the normal distribution or the Poisson distribution.

3.1 Central limit theorem (approximation with normal distribution)

If n is large and p is not too close to 0 or 1, the binomial distribution can be approximated by the normal distribution.

For $S_n \sim \text{Bin}(n, p)$; $E(S_n) = np$, $\text{Var}(S_n) = \sigma^2 = np(1 - p)$,

$$\lim_{n \rightarrow \infty} P\left(a \leq \frac{S_n - \mu}{\sigma} \leq b\right) = \int_a^b \varphi(x) dx = \Phi(b) - \Phi(a) \quad (34)$$

where φ is the standard normal pdf. This is the central limit theorem, which states that the binomial random variables approaches a normal distribution when $np(1 - p) > 10$.

3.1.1 continuity correction

$$P(a \leq S_n \leq b) = P(a - 0.5 \leq S_n \leq b + 0.5) \quad (35)$$

where $S \sim \text{Bin}(n, p)$ and a, b are integers. Useful when a, b are close, and $np(1 - p)$ is not large.

3.1.2 Law of large numbers

For

$$\begin{aligned} S_n \sim \text{Bin}(n, p) ; E(S_n) = np, E\left(\frac{S_n}{n}\right) &= p \\ P\left(\left|\frac{S_n}{n} - p\right| < \varepsilon\right) &\rightarrow 1 \text{ as } n \rightarrow \infty \end{aligned} \quad (36)$$

In English, this is saying that, as n is large, the frequency of success in n trials will converge to the probability of success p .

3.1.3 Confidence interval

In most cases, if real probability of success is unknown, we can use the Law of large number to

1. approximate p
2. find confidence interval $(\hat{p} - \varepsilon, \hat{p} + \varepsilon)$ (know how accurate the approximation is.) Connecting law of large number with CLT, we can proof that

$$P(|\hat{p} - p| < \varepsilon) \geq 2\Phi(2\varepsilon\sqrt{n}) - 1 \quad (37)$$

where, $2\Phi(2\varepsilon\sqrt{n}) - 1$ is the confidence level, i.e. how confident we are that the real probability is in the interval.

3.2 Poisson Distribution

3.2.1 Poisson r.v.

A discrete r.v. L has the Poisson distribution with parameter $\lambda > 0$ if its pmf is

$$p_L(k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad (38)$$

for $k = 0, 1, 2, \dots$

- write $L \sim \text{Poisson}(\lambda)$
- The mean and variance of a Poisson r.v. are both equal to λ .

3.2.2 Law of rare events

For $S_n \sim \text{Bin}\left(n, \frac{\lambda}{n}\right)$, where $\frac{\lambda}{n} < 1$, S_n follows the law of rare events,

$$\lim_{n \rightarrow \infty} P(S_n = k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad (39)$$

The distribution $\text{Bin}(n, \frac{\lambda}{n})$ approaches $\text{Poisson}(\lambda)$ distribution, where $E(S_n) = \lambda$

For a fixed n , to quantify the error in approximation, we have:

Let $X \sim \text{Bin}(n, p)$, and $Y \sim \text{Poisson}(\lambda)$, where $\lambda = np$

then for any subset

$$\begin{aligned} A &\subseteq \{0, 1, 2, \dots, n\}, k \in A \\ |P(X = k) - P(Y = k)| &\leq np^2 \end{aligned} \quad (40)$$

if $np^2 < 1$, then the approximation is good, and that

$$P(X = k) \approx P(Y = k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad (41)$$

3.3 Exponential Distribution

No mentioning where it comes from, but will be told when “can be modeled by exponential distribution” A continuous r.v. X has the exponential distribution with parameter $\lambda > 0$ if its pdf is

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (42)$$

Write $X \sim \text{Exp}(\lambda)$ The cdf is found by integrating the pdf,

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{for } x \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (43)$$

Expectations and variance are

$$E(X) = \frac{1}{\lambda}, \text{Var}(X) = \frac{1}{\lambda^2} \quad (44)$$

- EXp distribution is memoryless, i.e.

$$\begin{aligned}
P(X > t + s \mid X > t) &= \frac{P(X > t + s, X > t)}{P(X > t)} \\
&= \frac{P(X > t + s)}{P(X > t)} \\
&= \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} \\
&= e^{-\lambda s} \\
&= P(X > t)
\end{aligned} \tag{45}$$

for all $s, t > 0$

4 Joint Distribution

4.1 discrete joint distribution

- definition:

$$p(k_1, k_2, k_3) = P(X_1 = k_1, X_2 = k_2, X_3 = k_3) \tag{46}$$

for r.v. $X_1 = k_1, X_2 = k_2, X_3 = k_3$

- expectation:

$$E(g(X_1, X_2, X_3)) = \sum_{k_1} \sum_{k_2} \sum_{k_3} g(k_1, k_2, k_3) p(k_1, k_2, k_3) \tag{47}$$

- marginal distribution:

$$p_1(k) = \sum_{k_2} \sum_{k_3} p(k, k_2, k_3) \tag{48}$$

- Multinomial distribution when looking for the probability of some independent events together, we can use the multinomial distribution.

$$P(X_1 = k_1, X_2 = k_2, X_3 = k_3) = \frac{n!}{k_1! k_2! k_3!} p_1^{k_1} p_2^{k_2} p_3^{k_3} \tag{49}$$

abbreviate this as $(X_1, X_2, \dots, X_r) \sim \text{Multi}(n, r, p_1, p_2, \dots, p_r)$

4.2 Continuous joint distribution

- definition:

$$P((X_1, X_2, X_3) \in A) = \int_A f(x_1, x_2, x_3) dx_1 dx_2 dx_3 \tag{50}$$

for r.v. X_1, X_2, X_3 and set $A \in \mathfrak{R}$

- expectation:

$$E(g(X_1, X_2, X_3)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x_1, x_2, x_3) f(x_1, x_2, x_3) dx_1 dx_2 dx_3 \tag{51}$$

- marginal distribution:

$$f_1(x) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y, z) dy dz \tag{52}$$

4.3 Independent joint random variables

- Necessary and sufficient Condition:

- discrete

$$p(x_1, x_2) = p_1(x_1)p_2(x_2) \quad (53)$$

- Continuous

$$f(x_1, x_2) = f_1(x_1)f_2(x_2) \quad (54)$$

- If two r.v. depend on different parameters, they are independent. i.e.

$$\begin{aligned} Y = f(X_1, X_2, X_3); \quad Z = g(X_4, X_5, X_6) \\ \Rightarrow Y \text{ and } Z \text{ are independent} \end{aligned} \quad (55)$$