MATH 241 Results

Daniel Kim

December 2021

Proposition 0.1 (Markov's Inequality)
Let X be a non-negative random variable and $a > 0$. Then

$$P(X \geq a) \leq \frac{E(X)}{a}.$$

Proposition 0.2 (Chebyshev's Inequality)
Let X be an arbitrary random variable and $d > 0$. Then

$$P(|X - EX| \geq d) \leq \frac{\mathrm{Var}(X)}{d^2}.$$

Example 0.3 (Confidence Intervals)
Suppose $X \sim \mathrm{Bin}(n, p)$, and $\hat{p} = X/n$. Then for any p, $P(p \in [\hat{p} - \epsilon, \hat{p} + \epsilon]) \geq c$ means $[\hat{p} - \epsilon, \hat{p} + \epsilon]$ is a **confidence interval** with confidence level c. Also,

$$P(p \notin [\hat{p} - \epsilon, \hat{p} + \epsilon]) = P(|\hat{p} - p| > \epsilon) \leq \frac{\mathrm{Var}(\hat{p})}{\epsilon^2} = \frac{p(1-p)}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2}.$$

Definition 0.4. The **cumulative distribution function (CDF)** of a random variable X is $F_X(x) = P(X \leq x)$, so that $F_X : \mathbb{R} \to [0, 1]$.
Properties:

- F_X is non-decreasing
- $F(-\infty) = 0$, $F(\infty) = 1$
- Right-continuous: $F(x) \to F(a)$ as $x \to a^+$.

Definition 0.5. A random variable X is **continuous** if its CDF can be expressed as an integral, i.e. there exists a nonnegative function $f_X : \mathbb{R} \to \mathbb{R}$ such that

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt \ \forall x.$$

We call f_X the **probability density function (PDF)** of X, so that $f_X = F_X'$.
Properties:

- Non-negativity: $f_X(x) \geq 0$
- Normalization:

$$\int_{-\infty}^{\infty} f_X(x)\, dx = 1$$

Also, LOTUS for continuous random variables also works:

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

Definition 0.6. The **support** of X is a set consisting of all x where $f_X(x) > 0$.

Definition 0.7. A continuous random variable X is said to be **uniformly distributed** in the interval (a, b), denoted by $X \sim \text{Unif}(a, b)$, if it has the following PDF:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{else} \end{cases}.$$

Corollary 0.8

Let $X \sim \text{Unif}(a, b)$. Then

$$F_X(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x < b \\ 1 & x \geq b \end{cases}$$

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

$$\text{Var}(X) = \frac{(b-a)^2}{12}$$

Definition 0.9. A continuous random variable X is said to be **exponentially distributed** with parameter $\lambda > 0$, denoted by $X \sim \text{Expo}(\lambda)$, if it has the following PDF:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{else} \end{cases}.$$

Corollary 0.10

Let $X \sim \text{Expo}(\lambda)$. Then

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$$E(X) = \frac{1}{\lambda}$$

$$\text{Var}(X) = \frac{1}{\lambda^2}$$

Proposition 0.11 (Memoryless property of exponential distribution)

Let $X \sim \text{Expo}(\lambda)$. Then

$$P(X > s + t \mid X > t) = P(X > s), \;\; s, t > 0.$$

Definition 0.12. Let F be the CDF of a continuous distribution. Then F^{-1} is the **inverse CDF**, also called the **quantile function**.

Proposition 0.13

Let $U \sim \text{Unif}(0, 1)$. Then the CDF of $X = F^{-1}(U)$ is F. Conversely, given X with CDF F, $U = F(X)$ is distributed as $\text{Unif}(0, 1)$.

Corollary 0.14

Given X with CDF F, we can generate random variable Y with CDF G as $Y = G^{-1}(F(X))$.

Proposition 0.15

Let $Y = aX + b$, where $a \neq 0$. Then

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right).$$

$$E(aX + b) = aE(X) + b$$

$$\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$$

Definition 0.16. Let $E(X) = \mu$ and $\mathrm{SD}(x) = \sigma$. Then $Y = \frac{X-\mu}{\sigma}$ has zero mean and unit variance, called the **standardized version** of X.

Proposition 0.17 (Monotone transformation)

Let $Y = g(X)$, where g is monotone. Then

$$f_Y(y) = \frac{1}{|g'(g^{-1}(y))|} f_X(g^{-1}(y)).$$

Definition 0.18. A continuous random variable X is said to have the **standard normal (Gaussian) distribution** if it has the following PDF:

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = \varphi(x).$$

Then $E(X) = 0$, $\mathrm{Var}(X) = \mathrm{SD}(X) = 1$.
Also,

$$\Phi(x) = P(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt.$$

Definition 0.19. A continuous random variable X is said to have the **normal (Gaussian) distribution** with mean μ and variance σ^2 denoted by $X \sim N(\mu, \sigma^2)$ if it has the following PDF:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right).$$

Also,

$$F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

Remark 0.20. The standard normal distribution is the standardized version of the normal Gaussian distribution.

Theorem 0.21 (CLT for binomials)

For any b,

$$\lim_{n\to\infty} P\left(\frac{X - np}{\sqrt{npq}} \leq b\right) = \Phi(b) = \int_{-\infty}^{b} \varphi(x)\, dx = \int_{-\infty}^{b} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$$

In other words, X is approximately distributed as $\tilde{X} \sim N(np, npq)$

Definition 0.22. The **joint cumulative distribution function (CDF)** of random variables X and Y is

$$F_{XY}(x, y) = P(X \leq x, Y \leq y)$$

Definition 0.23. We call f_{XY} the **joint PDF** of (X, Y) given by

$$f_{XY}(x, y) = \frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y}$$

Properties:

- Non-negativity: $f_{XY}(x, y) \geq 0$.
- Normalization: $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y)\, dx dy = 1$.

Corollary 0.24
For region $A \subset \mathbb{R}^2$,

$$P((X, Y) \in A) = \iint_A f_{XY}(x, y)\, dx dy.$$

Corollary 0.25 (LOTUS in 2-D)

$$E(g(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{XY}(x, y)\, dx dy$$

Corollary 0.26
Extracting marginal CDF from joint CDF:

$$F_X(x) = F_{XY}(x, \infty),$$
$$F_Y(y) = F_{XY}(\infty, y).$$

Extracting marginal PDF from joint PDF:

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y)\, dy,$$
$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y)\, dx.$$

Definition 0.27. Let $A \subset \mathbb{R}^2$ be a region on the plane with finite area. We say (X, Y) is **uniformly distributed** over A if the joint PDF is

$$f_{XY}(x, y) = \begin{cases} \frac{1}{\text{area}(A)} & (x, y) \in A \\ 0 & \text{else} \end{cases}$$

Definition 0.28. A pair of continuous random variables X, Y are **continuous** if any of the following is true:

- $f_{XY}(x, y) = f_X(x) f_Y(y)$ for all x, y
- $F_{XY}(x, y) = F_X(x) F_Y(y)$

Definition 0.29. We denote

$$p_{Y|X}(y \mid x) = P(Y = x \mid X = x) = \frac{p_{XY}(x, y)}{p_X(x)}$$

as the **conditional PMF** of Y given $X = x$, provided that $p_X(x) > 0$.

Definition 0.30. For continuous (X, Y), the **conditional PDF** of Y given $X = x$ is defined as

$$f_{Y|X}(y \mid x) = \frac{f_{XY}(x, y)}{f_X(x)}$$

provided that $f_X(x) > 0$.

Corollary 0.31

If X and Y are independent, then $f_{Y|X}(y \mid x) = f_Y(y)$ for all x such that $f_X(x) > 0$.

Definition 0.32. $X_1, X_2, \ldots, X_n$ are **mutually independent** if any of the following is true:

- For any intervals $I_1, \ldots, I_n \subset \mathbb{R}$,

$$P(X_1 \in I_1, \ldots, X_n \in I_n) = P(X_1 \in I_1) \times \cdots \times P(X_n \in I_n).$$

- For any real-valued functions $f_1, \ldots, f_n$,

$$E(f_1(X_1) \times \cdots \times f_n(X_n)) = E(f_1(X_1)) \times \cdots \times E(f_n(X_n)).$$

Definition 0.33. Continuous random variables $X_1, X_2, \ldots, X_n$ are **mutually independent** if their joint PDF factorizes as a product of marginal PDFs:

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = f_{X_1}(x_1) \times \cdots \times f_{X_n}(x_n).$$

Definition 0.34. Let $S = X + Y$ where X, Y are independent continuous random variables.

CDF of S:

$$F_S(s) = P(X + Y \leq s) = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{s-x} f_{XY}(x, y)\, dy \right) dx$$

PDF of S:

$$f_S(s) = F_S'(s) = \int_{-\infty}^{\infty} f_X(x) f_Y(s - x)\, dx$$

This operation is known as **convolution**, denoted by $f_S = f_X * f_Y$.

Corollary 0.35

Let $X_1, \ldots, X_n$ be independent normals with $X_i \sim N(\mu_i, \sigma_i^2)$, then $S = X_1 + \cdots + X_n \sim N(\mu_1 + \cdots + \mu_n, \sigma_1^2 + \cdots + \sigma_n^2)$.

Furthermore, if $X \sim N(\mu, \sigma^2)$, then $aX + b \sim N(a\mu + b, a^2\sigma^2)$.

Theorem 0.36 (Cramér's Theorem)

If X and Y are independent and $X + Y$ is normal, then both X and Y must be normal.

Proposition 0.37 (2-D Transformation)

Suppose

$$\begin{pmatrix} U \\ V \end{pmatrix} = A \begin{pmatrix} X \\ Y \end{pmatrix}$$

for matrix A representing a one-to-one and differentiable mapping $\mathbb{R}^2 \to \mathbb{R}^2$. Then

$$f_{UV}(u, v) = \frac{1}{|\det(A)|} f_{XY}(A^{-1}(u, v)).$$

Proposition 0.38
Suppose

$$\begin{pmatrix} du \\ dv \end{pmatrix} = \begin{pmatrix} \partial u/\partial x & \partial u/\partial y \\ \partial v/\partial x & \partial v/\partial y \end{pmatrix} \begin{pmatrix} dx \\ dy \end{pmatrix}$$

for Jacobian $\frac{\partial(u,v)}{\partial(x,y)}$ representing a one-to-one and differentiable mapping $\mathbb{R}^2 \to \mathbb{R}^2$. Then

$$f_{UV}(u,v) = \frac{f_{XY}(g^{-1}(u,v))}{\left|\det \frac{\partial(u,v)}{\partial(x,y)}\right|} = \left|\det \frac{\partial(x,y)}{\partial(u,v)}\right| f_{XY}(g^{-1}(u,v)).$$

Definition 0.39. The **covariance** of random variables X and Y is

$$\text{Cov}(X,Y) = E((X-EX)(Y-EY)) = E(XY) - (EX)(EY).$$

Properties:

$$\text{Cov}(X,X) = \text{Var}(X)$$
$$\text{Cov}(aX+b, cY+d) = ac\text{Cov}(X,Y)$$
$$\rho(aX+b, cY+d) = \rho(X,Y),\ \ a,c>0$$
$$\text{Cov}(X+Y, W+Z) = \text{Cov}(X,W) + \text{Cov}(Y,W) + \text{Cov}(X,Z) + \text{Cov}(Y,Z)$$
$$\text{Cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j\right) = \sum_{i=1}^n \sum_{j=1}^m \text{Cov}(X_i, X_j)$$
$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y)$$
$$\text{Var}(X_1 + \cdots + X_n) = \text{Var}(X_1) + \cdots + \text{Var}(X_n) + \sum_{i \neq j} \text{Cov}(X_i, X_j)$$

Definition 0.40. The **correlation coefficient** of random variables X and Y is

$$\text{Corr}(X,Y) = \rho(X,Y) = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

Definition 0.41. We say ρ is **positively correlated** if $\rho > 0$. We say ρ is **negatively correlated** if $\rho < 0$.

Definition 0.42. We say X and Y are **uncorrelated** if $\text{Cov}(X,Y) = 0 \iff \rho(X,Y) = 0 \iff E(XY) = (EX)(EY)$.

Corollary 0.43
Independent variables are uncorrelated. However, the converse is false: uncorrelated variables are not necessarily independent.

Corollary 0.44
$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$ if and only if X and Y are uncorrelated.

Definition 0.45. The **mean squared error (MSE)** is $E[(Y-\hat{Y})^2]$ where $\hat{Y}$ is an estimate of unobserved random variable Y.

Definition 0.46. The **conditional expectation** of Y given $X = x$ is

$$E(Y \mid X = x) = \int y f_{Y|X}(y \mid x)\, dy$$

which is a function of x.

Definition 0.47. The **conditional variance** of Y given $X = x$ is

$$\mathrm{Var}(Y \mid X = x) = E(Y^2 \mid X = x) - E(Y \mid X = x)^2 = \int y^2 f_{Y|X}(y \mid x)\,dy - E(Y \mid X = x)^2$$

which is a function of x.

Proposition 0.48

Without observing X, the PDF of Y is $f_Y(y)$ and the best estimate is the unconditional mean $E(Y)$. This achieves MSE $= \mathrm{Var}(Y)$.

Upon observing $X = x$, the PDF of Y becomes $f_{Y|X}(y \mid x)$ and the best estimate is the conditional mean $E(Y \mid X = x)$. This achieves MSE $= E(\mathrm{Var}(Y \mid X)) = \mathrm{Var}(Y) - \mathrm{Var}(E(Y \mid X))$.

Theorem 0.49 (Law of total expectation)

Expectation of conditional mean = unconditional mean

$$E(E(Y \mid X)) = E(Y)$$

Proposition 0.50

Best linear estimate of Y given X is

$$\mu_Y + \frac{X - \mu_X}{\sigma_X}\sigma_Y \rho(X, Y)$$

where the linear estimator is useful if X and Y are correlated i.e. $\rho \neq 0$.

Theorem 0.51 (Cauchy-Schwarz Inequality)

For any random variables U, V, $(E(UV))^2 \leq E(U^2)E(V^2)$, with equality if and only if $U = cV$ for some constant c.

Theorem 0.52 (Law of Large Numbers)

Let $X_1, X_2, \ldots$ be a sequence of uncorrelated and identically distributed (iid) random variables with mean μ and variance σ^2. Let $S_n = X_1 + \cdots + X_n$ and $\overline{X}_n = \frac{S_n}{n}$. For any $\epsilon > 0$,

$$\lim_{n \to \infty} P(\left|\overline{X}_n - \mu\right| > \epsilon) = 0.$$

Remark 0.53. LLN only requires uncorrelatedness while CLT requires independence.

Definition 0.54. The **moment generating function (MGF)** of a random variable X is defined as

$$M_X(t) = E(e^{tX}) = \sum_{k \geq 0} \frac{t^k}{k!} E(X^k),$$

which is a function of $t \in \mathbb{R}$.

The kth moment of X is $E(X^k) = M_X^{(k)}(0)$ i.e. the kth derivative of the MGF at 0.

Corollary 0.55

For any constant a, b,

$$M_{aX+b}(t) = M_X(at)e^{bt}$$

Corollary 0.56

Let X and Y be independent. Then $M_{X+Y}(t) = M_X(t)M_Y(t)$.

Remark 0.57. This is helpful for turning convolutions into products.

Example 0.58

For $X \sim \text{Bern}(p)$, $M_X(t) = E(e^{tX}) = (1-p)\cdot e^0 + p \cdot e^t = 1 - p + pe^t$.

Example 0.59

For $X \sim N(0,1)$, $M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} e^{tx}\, dx = e^{t^2/2}$.

Corollary 0.60

Let $X_1, \dots, X_n$ be iid and $S_n = X_1 + \cdots + X_n$. Then $M_{S_n}(t) = (M_{X_1}(t))^n$.

Theorem 0.61 (Central Limit Theorem)

Let $X_1, X_2, \dots$ be iid with mean μ and variance σ^2. Let $S_n = X_1 + \cdots + X_n$. Then $\frac{S_n - n\mu}{\sqrt{n\sigma^2}}$ is approximately standard normal (in the sense of CDF):

$$\lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \le x\right) = \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt, \ \forall x \in \mathbb{R}.$$

Remark 0.62. Note that $\frac{S_n - n\mu}{\sqrt{n\sigma^2}}$ is the standardized version of $\overline{X_n}$ with mean μ and variance $\frac{\sigma^2}{n}$.

Remark 0.63. Essentially, the CLT says $\overline{X}_n \approx \mu + N\left(0, \frac{\sigma^2}{n}\right)$ i.e. $\overline{X}_n \approx N\left(\mu, \frac{\sigma^2}{n}\right)$ (the error term is approximately Gaussian like), while the LLN only tells us that $\overline{X}_n = \mu +$ some small error (we don't know how fast it vanishes as n grows).

Remark 0.64. CLT says that S_n is approximately distributed as $N(n\mu, n\sigma^2)$, and $\overline{X}_i = \frac{1}{n}S_n$ is approximately distributed as $N(\mu, \frac{\sigma^2}{n})$.