MATH 395 Notes

Contents

1 Joint Distributions

1.1 Joint Probability Mass Function

Let X and Y be discrete random variables. The joint probability mass function of X and Y is

$$p_{X,Y}(x,y) = \mathbb{P}(X = x, Y = y)$$

1.2 Joint Probability Density Function

Random variables X and Y are jointly continuous if there exists a joint density function $f_{X,Y}$ on $\mathbb{R}^2$ such that for subsets $B \subseteq \mathbb{R}^2$

$$f_{X,Y}(x,y) = \iint_B f(x,y)\ dy\ dx$$

1.3 Joint Cumulative Distribution Function

The cumulative distribution of random variables X and Y is

$$F(x,y) = \mathbb{P}(X \le x, Y \le y)$$

1.4 Joint Distributions of Independent Random Variables

Let X and Y be discrete random variables. X and Y are independent if and only if

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x) \cdot \mathbb{P}(Y = y)$$

Let X and Y be jointly continuous random variables. X and Y are independent if and only if

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$$

1.5 Marginal Probability Mass Function

Let X and Y be discrete random variables with joint probability mass function $p_{X,Y}(x,y)$. The marginal probability mass functions of X and Y are

$$p_X(x) = \sum_{y \in \Omega_Y} p_{X,Y}(x,y) \qquad p_Y(y) = \sum_{x \in \Omega_X} p_{X,Y}(x,y)$$

1.6 Marginal Probability Density Function

Let X and Y be continuous random variables with joint probability density function $f_{X,Y}(x,y)$. The marginal probability density functions of X and Y are

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dy \qquad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dx$$

1.7 Summary of Joint Distributions

	Discrete Random Variables	Continuous Random Variables
Joint PMF/PDF	$p_{X,Y}(x,y) = \mathbb{P}(X=x, Y=y)$	$f_{X,Y} \neq \mathbb{P}(X=x, Y=y)$
Joint range/support	$\{(x,y) \in \Omega_X \times \Omega_Y \mid p_{X,Y}(x,y) > 0\}$	$\{(x,y) \in \Omega_X \times \Omega_Y \mid f_{X,Y}(x,y) > 0\}$
Joint CDF	$F_{X,Y}(x,y) = \sum_{t \leq x} \sum_{s \leq y} p_{X,Y}(t,s)$	$F_{X,Y}(s,t) = \int_{-\infty}^{s} \int_{-\infty}^{t} f_{X,Y}(x,y)\, dy\, dx$
Normalization	$\sum_{x,y} p_{X,Y}(x,y) = 1$	$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy\, dx = 1$
Marginal PMF/PDF	$p_X(x) = \sum_{y \in \Omega_y} p_{X,Y}(x,y)$	$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy$
Expectation	$\mathbb{E}[g(X,Y)] = \sum_{x,y} g(x,y) \cdot p_{X,Y}(x,y)$	$\mathbb{E}[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) \cdot f_{X,Y}(x,y)\, dy\, dx$

1.8 Joint Uniform Random Variables

A joint uniform random variable (X,Y) is equally likely to be anywhere in a subset D of the Euclidean plane $\mathbb{R}^2$

- $f_{X,Y}(x,y) = \begin{cases} \frac{1}{\text{area}(D)} & \text{if (x, y)} \in \text{D} \\ 0 & \text{otherwise} \end{cases}$

A joint uniform random variable (X,Y,Z) is equally likely to be anywhere in a subset D of the Euclidean space $\mathbb{R}^3$

- $f_{X,Y,Z}(x,y,z) = \begin{cases} \frac{1}{\text{vol}(B)} & \text{if (x, y, z)} \in \text{D} \\ 0 & \text{otherwise} \end{cases}$

1.9 Multinomial Random Variables

A multinomial random variable $(X_1, ..., X_k)$ is the number of successes for each side of a k-sided dice rolled n times, where each side i has probability p_i, denoted $(X_1, ..., X_k) \sim \text{Mult}(n, k, p_1, ..., p_k)$

- $\mathbb{P}(X_1 = x_1, ..., X_k = x_k) = \binom{n}{x_1, ..., x_k} {p_1}^{x_1} ... {p_k}^{x_k}$
- $\mathbb{E}[X_i] = np_i$
- $\text{Var}(X_i) = np_i(1-p_i)$

1.10 Standard Bivariate Normal Distribution

The standard bivariate normal distribution is the joint probability density function for two possibly dependent standard normal random variables with covariance ρ

- $f_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{x^2+y^2-2\rho xy}{2(1-\rho^2)}}$
- X and Y are independent if and only if $\rho = 0$

2 Sums

2.1 Sums of Independent Random Variables

If X and Y are independent discrete random variables with probability mass functions p_X and p_Y, then the probability mass function of X and Y is

$$p_{X+Y}(z) = \sum_k p_X(k)p_Y(z-k) = \sum_k p_X(z-k)p_Y(k)$$

If X and Y are independent continuous random variables with probability density functions f_X and f_Y, then the probability density function of X and Y is

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)\ dx = \int_{-\infty}^{\infty} f_X(z-x)f_Y(x)\ dx$$

2.2 Sums of Independent Moment Generating Functions

If X and Y are independent random variables with moment generating functions $M_X(t)$ and $M_Y(t)$, then

$$M_{X+Y}(t) = M_X(t)M_Y(t)$$

2.3 Sums of Independent Poisson Random Variables

Let $X \sim \mathrm{Poi}(\lambda_1)$ and $Y \sim \mathrm{Poi}(\lambda_2)$ such that $\lambda = \lambda_1 + \lambda_2$. Let $Z = X + Y$. Then $Z \sim \mathrm{Poi}(\lambda)$

- $\mathbb{P}(Z = z) = \dfrac{e^{-(\lambda_1+\lambda_2)}}{z!}(\lambda_1 + \lambda_2)^z$
- $\mathbb{E}[Z] = \lambda_1 + \lambda_2$
- $\mathrm{Var}(Z) = \lambda_1 + \lambda_2$

2.4 Sums of Independent Binomial Random Variables

Let $X \sim \mathrm{Bin}(n_1, p)$ and $Y \sim \mathrm{Bin}(n_2, p)$ such that $n = n_1 + n_2$. Let $Z = X + Y$. Then $Z \sim \mathrm{Bin}(n, p)$

- $\mathbb{P}(Z = z) = \binom{n_1+n_2}{z} p^z (1-p)^{(n_1+n_2)-z}$
- $\mathbb{E}[Z] = (n_1 + n_2)p$
- $\mathrm{Var}(X) = (n_1 + n_2)(1-p)p$

2.5 Sums of Independent Normals

Let $S_n = X_1 + ... + X_n$, where $X_1, ..., X_n$ are independent and identically distributed (iid) random variables each with expectation μ and variance σ^2

- $\mathbb{E}[S_n] = n\mu$
- $\mathrm{Var}(S_n) = n\sigma^2$

2.6 Negative Binomial Random Variables

A negative binomial random variable X models the number of independent trials $Y_i \sim \mathrm{Ber}(p)$ before seeing the r^{th} success. $X = \sum_{i=1}^{r} Z_i$ where $Z_i \sim \mathrm{Geo}(p)$, denoted $X \sim \mathrm{NegBin}(r, p)$

- $\mathbb{P}(X = x) = \binom{x-1}{r-1}(1-p)^{x-r}p^r$
- $\mathbb{E}[X] = \frac{r}{p}$
- $\mathrm{Var}(X) = \frac{r(1-p)}{p^2}$

The negative binomial random variable models the sum of independent geometric random variables

2.7 Gamma Distribution

A gamma distribution X models the sum of n independent exponential random variables $Y_i \sim \mathrm{Exp}(\lambda)$ where $X = \sum_{i=1}^{n} Y_i$, denoted $X \sim \mathrm{Gamma}(n, \lambda)$

- $f_X(x) = \dfrac{\lambda^n x^{n-1}}{\displaystyle\int_0^\infty x^{n-1}e^{-x}\,dx}e^{-\lambda x}$
- $\mathbb{E}[X] = \frac{n}{\lambda}$
- $\mathrm{Var}(X) = \frac{n}{\lambda^2}$

3 Symmetry

3.1 Symmetric Functions

A function $f : \mathbb{R}^n \to \mathbb{R}$ is symmetric if $f(x_{\sigma_1}, x_{\sigma_2}, ..., x_{\sigma_n}) = f(x_1, x_2, ..., x_n)$ for any permutation $(\sigma_1, \sigma_2, ..., \sigma_n)$ of $(1, 2, ..., n)$

3.2 Exchangeable Random Variables

A sequence of random variables $X_1, X_2, ..., X_n$ is exchangeable if $(X_1, X_2, ..., X_n) \stackrel{d}{=} (X_{\sigma_1}, X_{\sigma_2}, ..., X_{\sigma_n})$

- If $X_1, X_2, ..., X_n$ are discrete random variables with joint probability mass function p, then these random variables are exchangeable if and only if p is a symmetric function
- If $X_1, X_2, ..., X_n$ are jointly continuous random variables with joint probability density function f, then these random variables are exchangeable if and only if f is a symmetric function
- If $X_1, X_2, ..., X_n$ are exchangeable, then these random variables are identically distributed such that they have the same marginal density function
- If $X_1, X_2, ..., X_n$ are independent and identically distributed (iid), then these random variables are exchangeable

3.3 Exchangeability in Sampling Without Replacement

Let $X_1, ..., X_k$ denote the outcomes of successive draws without replacement from the set $\{1, ..., n\}$. Then $X_1, ..., X_k$ are exchangeable

4 Multivariate Expectation and Variance

4.1 Linearity of Expectation

Let X and Y be random variables with finite expectations and let $a, b \in \mathbb{R}$. Then

$$\mathbb{E}[aX + bY + c] = a\mathbb{E}[X] + b\mathbb{E}[Y] + c$$

- Linearity of expectation applies for both independent and dependent variables

4.2 Product of Expectations

Let X and Y be independent random variables. Then for all functions g_X and g_Y

$$\mathbb{E}\left[g_X(X) \cdot g_Y(Y)\right] = \mathbb{E}[g_X(X)] \cdot \mathbb{E}[g_Y(Y)]$$

4.3 Sums of Independent Variances

Let X and Y be independent random variables with finite variances and let $a, b \in \mathbb{R}$. Then

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

5 Covariance and Correlation

5.1 Covariance

The covariance of two random variables X and Y is

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])] \\ &= \mathbb{E}[XY]-\mathbb{E}[X]\mathbb{E}[Y]\end{aligned}$$

- If X and Y are independent, then $\mathrm{Cov}(X,Y)=0$
- If $X=Y$, then $\mathrm{Cov}(X,Y)=\mathrm{Var}(X)=\mathrm{Var}(Y)$

Let $X_1,..,X_n$ and $Y_1,...,Y_m$ be random variables and let $a_1,...,a_n,b_1,...,b_m\in\mathbb{R}$. Then

$$\mathrm{Cov}\left(\sum_{i=1}^{n}a_iX_i,\sum_{j=1}^{m}b_jY_j\right)=\sum_{i=1}^{n}\sum_{j=1}^{m}a_ib_j\mathrm{Cov}(X_i,X_j)$$

5.2 Covariance Matrix

The covariance matrix of random variables X,Y is defined as

$$S=\begin{pmatrix}\mathrm{Cov}(X,X) & \mathrm{Cov}(X,Y)\\ \mathrm{Cov}(Y,X) & \mathrm{Cov}(Y,Y)\end{pmatrix}\quad S^{-1}=\begin{pmatrix}\dfrac{1}{\mathrm{Var}(X)} & -\dfrac{\mathrm{Corr}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}\\ -\dfrac{\mathrm{Corr}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}} & \dfrac{1}{\mathrm{Var}(Y)}\end{pmatrix}$$

5.3 Sums of Variances

Let $X_1,...,X_n$ be random variables with finite variances and covariances. Then

$$\mathrm{Var}\left(\sum_{i=1}^{n}X_i\right)=\sum_{i=1}^{n}\mathrm{Var}(X_i)+2\sum_{1\le i<j\le n}\mathrm{Cov}(X_i,X_j)$$

5.4 Uncorrelated Random Variables

The random variables X and Y are uncorrelated if $\mathrm{Cov}(X,Y)=0$

- If X and Y are independent, then they are uncorrelated
  - The converse does not hold

5.5 Correlation

The correlation of two random variables X and Y is

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}$$

- $-1 \leq \mathrm{Corr}(X,Y) \leq 1$
- Let $a, b \in \mathbb{R}$ with $a \neq 0$. Then $\mathrm{Corr}(aX+b, Y) = \frac{a}{|a|}\mathrm{Corr}(X,Y)$
- $\mathrm{Corr}(X,Y) = 1$ if and only if there exists $a > 0$ and $b \in \mathbb{R}$ such that $Y = aX + b$
- $\mathrm{Corr}(X,Y) = -1$ if and only if there exists $a < 0$ and $b \in \mathbb{R}$ such that $Y = aX + b$

6 Tail Bounds and Limit Theorems

6.1 Markov's Inequality

Let X be a non-negative random variable. Then for any $c > 0$

$$\mathbb{P}(X \geq c) \leq \frac{\mathbb{E}[X]}{c}$$

6.2 Chebyshev's Inequality

Let X be a random variable with a finite mean and finite variance. Then for any $c > 0$

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq c) \leq \frac{\mathrm{Var}(X)}{c^2}$$

- $\mathbb{P}(X \geq \mathbb{E}[X] + c) \leq \dfrac{\mathrm{Var}(X)}{c^2}$
- $\mathbb{P}(X \leq \mathbb{E}[X] - c) \leq \dfrac{\mathrm{Var}(X)}{c^2}$

6.3 Law of Large Numbers With Finite Variances

Let $X_1, X_2, X_3, ...$ be independent and identically distributed random variables with finite mean and variance and let $S_n = X_1 + ... + X_n$. Then for any fixed $\varepsilon > 0$

$$\lim_{n \to \infty} \mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}[X_1]\right| < \varepsilon\right) = 1$$

6.4 Central Limit Theorem

Let $X_1, X_2, X_3, ...$ be independent and identically distributed random variables with finite mean μ and variance σ^2 and let $S_n = X_1 + ... + X_n$

- $\mathbb{E}[S_n] = n\mu$
- $\mathrm{Var}(S_n) = n\sigma^2$

The CDF of $Y_n = \dfrac{S_n - n\mu}{\sigma\sqrt{n}}$ converges to the CDF of the standard unit normal $\mathcal{N}(0, 1)$

- $\mathbb{E}[Y_n] = 0$
- $\mathrm{Var}(Y_n) = 1$

Alternately, the CDF of $\bar{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ converges to the CDF of normal variable $\mathcal{N}\left(\mu, \dfrac{\sigma^2}{n}\right)$

- $\mathbb{E}\left[\bar{X}\right] = \mu$
- $\mathrm{Var}\left(\bar{X}\right) = \dfrac{\sigma^2}{n}$

7 Conditional Distributions

7.1 Conditional Probability Mass Function

Let X be a discrete random variable. Then the conditional probability mass function of X given event $Y = y$ is

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

7.2 Conditional Probability Density Function

Let X and Y be jointly continuous random variables with joint density function $f_{X,Y}(x, y)$. Then the conditional probability density function of X given event $Y = y$ is

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

7.3 Conditional Expectation for Discrete Variables

Let X be a discrete random variable. Then the conditional expectation of X given event $Y = y$ is

$$\mathbb{E}[X \mid Y = y] = \sum_{x \in \Omega_X} x \cdot p_{X|Y}(x \mid y) = \sum_{x \in \Omega_X} x \cdot \mathbb{P}(X = x \mid Y = y)$$

7.4 Conditional Expectation for Continuous Variables

Let X be a continuous random variable. Then the conditional expectation of X given event $Y = y$ is

$$\mathbb{E}[X \mid Y = y] = \int_{-\infty}^{\infty} x \cdot f_{X|Y}(x \mid y)\, dx$$

7.5 Law of Total Probability for Discrete Variables

Let X and Y be discrete random variables. Then the probability mass function of X is

$$\mathbb{P}(X = x) = \sum_{y \in \Omega_Y} p_{X|Y}(x \mid y) \cdot \mathbb{P}(Y = y) = \sum_{y \in \Omega_Y} \mathbb{P}(X = x \mid Y = y) \cdot \mathbb{P}(Y = y)$$

7.6 Law of Total Probability for Continuous Variables

Let X and Y be continuous random variables. Then the probability density function of X is

$$f_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x \mid y) \cdot f_Y(y)\, dy$$

7.7 Law of Total Expectation for Discrete Variables

Let X and Y be discrete random variables. Then the expectation of X is

$$\mathbb{E}[X] = \sum_{y \in \Omega_Y} \mathbb{E}[X \mid Y = y] \cdot \mathbb{P}(Y = y)$$

7.8 Law of Total Expectation for Continuous Variables

Let X and Y be continuous random variables. Then the expectation of X is

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} \mathbb{E}[X \mid Y = y] \cdot f_Y(y)\, dy$$

7.9 Wald's Identity

Let $X_1, X_2, X_3, ...$ be independent and identically distributed random variables with finite mean. Let N be a non-negative integer-valued variable independent of $X_1, X_2, X_3, ...$ also with finite mean. Let $S_N = X_1 + ... + X_N$. Then

$$\mathbb{E}[S_N] = \mathbb{E}[N] \cdot \mathbb{E}[X_1]$$