1 Chapter 1

The **union** of two events $C = A \cup B$ is the event C where either A occurs or B occurs or both occur

The **intersection** of two events $C = A \cap B$ is the event C where both A and B occurs

The **complement** of an events A is A^c is the event that A does not occur

Commutative Laws:

- $A \cup B = B \cup A$
- $A \cap B = B \cap A$

Associative Laws:

- $(A \cup B) \cup C = A \cup (B \cup C)$
- $(A \cap B) \cap C = A \cap (B \cap C)$

Distributive Laws:

- $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$
- $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$

Propability measures For mutually disjoint events $A_1, A_2, \ldots, A_n$

$P\left(\cup_{i=1}^{\infty} A_i\right) = \Sigma_{i=1}^{\infty} P\left(A_i\right)$

Property A	$P\left(A^c\right) = 1 - P\left(A\right)$
Property B	$P\left(\emptyset\right) = 0$
Property C	If $A \subset B$ Then $P\left(A\right) \leq P\left(B\right)$
Property D	$P\left(A \cup B\right) = P\left(A\right) + P\left(B\right) - P\left(A \cap B\right)$, addition law

$\binom{n}{r} = \frac{n!}{(n-r)!r!}$

Proposition C: The number of ways that n objects can be grouped into r calsses with n_i in the ith class, $i = 1, \ldots, r$ and $\Sigma_{i=1}^{r} n_i = n$ is $\binom{n}{n_1 n_2 \ldots n_r} = \frac{n!}{n_1! n_2! \ldots n_r!}$. This can be extented to multinomial coefficients. $(x_1 + x_2 + \ldots x_r)^n = \Sigma \binom{n}{n_1 n_2 \ldots n_r} x_1^{n_1} x_2^{n_2} \ldots x_r^{n_r}$

Condiitonal Probability: Let A and B be two events with $P\left(B\right) \neq 0$. The conditional probability of A given B is defined to be $P\left(A \mid B\right) = \frac{P(A \cap B)}{P(B)}$

Multiplication Law: Let A and B be events and assume $P\left(B\right) \neq 0$ Then $P\left(A \cap B\right) = P\left(A \mid B\right) P\left(B\right)$

Law of Total Probability

Let $B_1, B_2, \ldots, B_n$ be such that $\cup_{i=1}^{n} B_i = \Omega$ and $B_i \cap B_j = \emptyset$ for $i \neq j$ all with non zero probablity. Then for any event A: $P\left(A\right) = \Sigma_{i=1}^{n} P\left(A \mid B_i\right) P\left(B_i\right)$

Bayes Rule: Let A and $B_1, \ldots, B_n$ where the events B_i are disjoint, and non-zero. Then $P\left(B_j \mid A\right) = \frac{P(A|B_j)P(B_j)}{\Sigma_{i=1}^{n} P(A|B_i)P(B_i)}$

Independence. If two events are independent then: $P\left(A\right) = P\left(A \mid B\right) = \frac{P(A \cap B)}{P(B)}$ then $P\left(A \cap B\right) = P\left(A\right) P\left(B\right)$

2 Chapter 2

Discrete Random Variable is a random variable that can take only a finite or at most a countably finite number of values.

There is a function p such that $P\left(x_i\right) = P\left(X = x_i\right)$ and $\Sigma_i p\left(x_i\right) = 1$. This function is called the probability mass function or the frequency function of the random variable X.

Cumulative Distribution Function of a random variable $F\left(x\right) = P\left(X \leq x\right)$, $-\infty < x < \infty$. The function is non decreasing $\lim_{x \to -\infty} F\left(x\right) = 0$ and $\lim_{x \to \infty} F\left(x\right) = 1$

Bernouli Random Variable: takes 2 values either 0 or 1 with probabilities $1 - p$ and p respectively.

$$\begin{array}{ll} P\left(1\right) = p & a \\ P\left(1\right) = p & a \\ P\left(1\right) = p & if x \neq 0 and x \neq 1 \end{array}$$

$$p(x) = \begin{cases} p^x (1-p)^{1-x}, & \text{if} x = 0 \text{ or } x = 1 \\ 0, & \text{otherwise} \end{cases}$$

$E(X) = p$ and $var(X) = p(1-p)$

Indicator Random Vairable I_A takes the value 1 if A occures and 0 otherwise. I_A is a Bernouilli random variable.

$$I_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A \\ 0, & \text{otherwise} \end{cases}$$

Binomial Distribution: n experimets, k successes with probability p,

$$p(k) = \binom{n}{k} p^k \left(1-p\right)^{n-k}$$

$E(X) = np$ and $var(X) = np(1-p)$

Geometric distribution: with successes probability p, probability first success at attempt X. $X = k$, so we have $k-1$ failures. For $k = 1, 2, \ldots$

$$p\left(k\right) = P\left(X = k\right) = \left(1-p\right)^{k-1} p$$

$E(X) = \frac{1}{p}$ and $var(X) = \frac{1-p}{p^2}$

Negative Geometric distribution: with successes probability p, let X denote total trials. how many tries to get r successes.

$$p(X = k) = \binom{k-1}{r-1} p^r \left(1-p\right)^{k-r}$$

$E(X) = \frac{k}{p}$ and $var(X) = \frac{(1-p)k}{p^2}$

Hypergeometric distribution: If a bag has n balls, of which r are black and $n - r$ are white.

Let X denote the number of black balls drawn then taking m balls without replacement

$$p(X=k)=\frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}}$$

$E(X)=\frac{mr}{n}$ and $var(X)=\frac{mr(n-r)(n-m)}{n^2(n-1)}$

Poisson frequency function: with parameter $\lambda, \lambda>0$

$$p(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$$

$e^{\lambda}=\Sigma_{k=0}^{\infty}\left(\frac{\lambda^k}{k!}\right)$

The **Poisson distribution** can be derived as the limit of a binomial distribution as the number of trials n approached infinity. and the probability of success on each trial p approaches zero such that $np=\lambda$

$$p(k)=\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}$$

$E(X)=\lambda$ and $var(X)=\lambda$

Continuous Random Variable: can take on a continuum of values, not a set value. The probability X lies between a and b is the areas under the curve. $P(a<X<b)=\int_a^b f(x)dx$, and $\int_{-\infty}^{\infty} f(x)dx=1$. $P(a<X<b)=P(a\leq X<b)=P(a<X\leq b)$

For small δ if f is continuous at x, $P(x-\frac{\delta}{2}\leq X\leq x+\frac{\delta}{2})=\int_{x-\frac{\delta}{2}}^{x+\frac{\delta}{2}} f(u)du\approx\delta f(x)$

The **cumulative distribution function** of a continuous random variable is $F(x)=P(X\leq x)$, where $F(x)=\int_{-\infty}^{x} f(u)du$. From calculus $f(x)=F'(x)$

Uniform distribution, To pick value over inter $[a,b]$, denoted $X\ U(a,b)$. the pdf:

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

$E(X)=\frac{a+b}{2}$ and $var(x)=\frac{(b-a)^2}{12}$

The cumulative distribution function of uniform distribition: $F_X(x)=\int_{-\infty}^{x} f_X(y)dy=\begin{cases}0, & \text{if } x<a\\ \frac{x-a}{b-a}, & \text{for } a\leq x\leq b\\ 1, & \text{if } x\geq b\end{cases}$

The Exponential Density

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

$E(X)=\frac{1}{\lambda}$ and $var(x)=\frac{1}{\lambda^2}$

The cumulative distribution function of exponential distribution: $F_X(x)=\int_{-\infty}^{x} f_X(y)dy=\begin{cases}\lambda e^{-\lambda x} & \text{for } x>0\\ 0, & \text{for } x\leq 0\end{cases}$

Memoryless Property: The exponential distribution satisfy the following property: $P(X>s+t|X>s)=P(X>t)$

The Gamma Density Function: depending on α, shape patameter, and λ, scale parameter. If $t<0, g(t)=0$

$$g(t)=\frac{\lambda^a}{\Gamma(\alpha)}e^{\alpha-1}r^{-\lambda t}, t\geq 0$$

$$\Gamma(x)=\int_0^{\infty} u^{x-1}e^{-u}du, x>0$$

By integration by parts $\Gamma(a)=(a-1)\Gamma(a-1)$

$E(X)=\frac{\alpha}{\lambda}$ and $Var(x)=\frac{\alpha}{\lambda^2}$

Normal Distribution: based on Central limit theorem. Depending on μ, mean, and σ, the standard deviation. $X\ N\left(\mu,\sigma^2\right)$ $f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{e(x-\mu)^2/2\sigma^2}$

If $Z\ N(0,1)$ The probability density function $\phi(x)=\frac{1}{\sqrt{2\pi}}e^{-X^2/2}$ The cumulatif density function $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-y^2/2}dy$

$E(X)=\mu$ and $var(x)=\sigma^2$

Beta Density: The beda density is useful for modeling ranodm variables that are restricted to the interval $[0,1]$: $f(u)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}u^{a-1}(1-u)^{b-1}, 0\leq u\leq 1$

Functions as random Variables

Proposition A: If $X\ N(\mu,\sigma^2)$ and $Y=aX+b$, then $Y\ N(a\mu+b,a^2\sigma^2)$

Proposition B: Lex X be a continuous random variable with density $f(x)$ and let $Y=g(x)$ where g is differentiable, strictly monotonic function on some intertal I. Suppose that $f(x)=0$ if x is not in I. Then Y has the density function $f_Y(y)=f_X(g^{-1}(y)))\left|\frac{d}{dy}g^{-1}(y)\right|$. For y such that $y=g(x)$, for some x, and $f_Y(y)=0$ if $y\neq g(x)$ for any x in I.

Proposition C: Let $Z=F(X)$; then Z has a uniform distribution on $[0,1]$

Proposition D: Let U be uniform on $[0,1]$, and let $X=F^{-1}(U)$. Then the cdf of X is F.

3 Chapter 3

The joint behavior of two random variables X and Y, is determined by the cumulative distribution function $F(X,y)=P(X\leq x,Y\leq y)$. Regardleess of whether X and X are continuous or discrete. This can be extented to $P(x_1<X\leq x_2,y_1<Y\leq y_2)=F(x_2,y_2)-F(x_2,y_1)-F(x_1,y_2)+F(x_1,y_1)$

Discrete Random Variables. Suppose that X and Y are discrete random variables defined on the same sample space. Their **joint frequency function**, or joint probability mass function $p(x,y)$ is $p(x_i, y_j) = P(X = x_i, Y = y_j)$

To find the frequency function of Y, we simply sum down the appropiate column of the table. For this reason P_Y is called the marginal frequency funciton of Y. $P_X(x) = \Sigma_i p(x, y_i)$ is the marginal frequency function of X.

Continuous Random Variables: Suppose that X and Y are continuous random variables with a joing cdf, $F(x,y)$. Their **joint density function** is a pievewise continous funciton of two variables $f(x,y)$. The density funciton $f(x,y)$ is nonnegative and $\int_{\infty}^{\infty}\int_{\infty}^{\infty} f(x,y)dydx = 1$

For 2 dimentional set of A: $P((X,Y) \in A) = \int\int_A f(x,y)dydx$. In particular, if $A = \{(X,Y)|X \leq x \text{ and } Y \leq y\}$ $F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)dudv$. To generalize $f(x,y) = \frac{\partial^2}{\partial x \partial y}F(x,y)$

for small δ_x and δ_y, if f is a continuous at (x,y): $P(x \leq X \leq x + \delta_x, y \leq Y \leq y + \delta_y) = \int_x^{x+\delta_x}\int_y^{y+\delta_y} f(u,v)dvdu \approx f(x,y)\delta_x\delta_y$

The marginal cdf of X or F_X, is $F_X(x) = P(X \leq x) = \lim_{y\to\infty} F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{\infty} f(u,y)dydu$. From this, it follows tha tthe density function of X alone, known as the marginal density of X is $f_X(x) = F_X'(x) = \int_{-\infty}^{\infty} f(x,y)dy$ In the discrete case, the marginal frequency function was found by summing the joint frequency funciton over the other variable, in the continous case it is found by integration.

In someapplications, it is usefyl to analyze distributions that are uniform over some region of Space. For example, in the plane, the random point (X,Y) is uniform over a region R, if for any $A \subset R$ $P((X,Y) \in A) = \frac{|A|}{|R|}$

Independent Random Variables: Random Variables $X_1, X_2, \dots X_n$ are said to be independent if therir joint cdf factors into the product of their marginal cfds: $F(x_1, x_2, \dots x_n) = F_{X_1}(x_1)F_{X_2}(x_2)\dots F_{X_n}(x_n)$ for all $x_1, x_2, \dots x_n$

Independence holds for both continuous and discrete random variables. $F(x,y) = F_X(x)F_Y(y)$ as well as $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$

Conditional distrbution.

For discrete case: $P(X = x_i|Y = y_j) = \frac{P(X=x_i,Y=y_j)}{P(Y=y_j)} = \frac{P_{XY}(x_i,y_j)}{P_Y(y_j)}$ Thus consequently $P_{XY}(x,y) = P_{X|Y}(x|y)P_Y(y)$ and $P_X(x) = \Sigma_y P_{X|Y}(x|y)\, P_Y(y)$

For continunous case: $f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)}$. We can then solve $P(y \leq Y \leq y + dy|x \leq X \leq X + dx) = \frac{f_{XY}(x,y)dxdy}{f_X(x)dx} = \frac{f_{XY}(x,y)}{f_X(x)}dy$ The joint density can be expressed in terms of the marginal and conditional densities: $f_{XY}(x,y) = f_{Y|X}(y|x)f_X(x)$, integrating both sides over x allows us to get the marginal density of Y to be expressed as $f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x)f_X(x)dx$

Suppose that X and Y are discrete random variables, and we have $Z = X+Y$. to get the frequency function of Z. $Z = z$, if $X = x$ and $Y = z - x$. $P_Z(z) = \Sigma_{x=-\infty}^{\infty} p(x, z-x)$ If X and Y are independent, $p(x,y) = P_X(x)P_Y(y)$ and $P_Z(z) = \Sigma_{x=-\infty}^{\infty} P_X(x)P_Y(z-x)$

To generalize $F_Z(z) = \int_{-\infty}^{\infty}\int_{-\infty}^{z} f(x, u - x)dudx = \int_{-\infty}^{z}\int_{-\infty}^{\infty} f(x, u-x)dxdu$ Which we can differentiate to get $f_Z(z) = \int_{-\infty}^{\infty} f(x, z-x)dx$

For the general case, suppose that X and Y are jointly distributed continuous random variables, that X and Y are mapped onto U and V by the transformation

$$u = g_1(x, y)$$
$$v = g_2(x, y)$$

and that the transformation can be inverted to obtain

$$x = h_1(u, v)$$
$$y = h_2(u, v)$$

Assume that g_1 and g_2 have continuous partial derivatives and that the Jacobian

$$J(x, y) = \det\begin{bmatrix} \frac{\partial g_1}{\partial x} & \frac{\partial g_1}{\partial y} \\ \frac{\partial g_2}{\partial x} & \frac{\partial g_2}{\partial y} \end{bmatrix} = \left(\frac{\partial g_1}{\partial x}\right)\left(\frac{\partial g_2}{\partial y}\right) - \left(\frac{\partial g_2}{\partial x}\right)\left(\frac{\partial g_1}{\partial y}\right) \neq 0$$

for all x and y. This leads directly to the following result.

Proposition A: Under the assumptions just stated, the joint density of U and V is: $f_{UV}(u,v) = \frac{F_{XY}(h_1(u,v),h_2(u,v))}{J(h_1(u,v),h_2(u,v))}$. For (u,v) such that $u = g_1(x,y)$ and $v = g_2(x,y)$ for some (x,y) and 0 elsewhere

Extrema and Order Statistics The density of $X_{(k)}$, the kth-order statistic, is $f_k(x) = \frac{n!}{(k-1)!(n-k)!}f(x)F^{k-1}(x)\left[1 - F(x)\right]^{n-k}$

4 Chapter 4

Expected Value, also known as mean. For discrete random variable X with frequency function $p(x)$, the expected value of X, denoted $E(X)$ is $E(X) = \Sigma_i x_i p(x_i)$ Provided the sum is less than infinity, if the sum diverges, the expectation in undefined.

For continuous ranodm variable with density $f(x)$, then $E(X) = \int_{-\infty}^{\infty} xf(x)dx$

If X is a nonnegative continuous random variable $E(X) = \int_0^{\infty}\left[1 - F(X)\right]dx$

markov's inequality: if X is a ranodm variable with $P(X \geq 0) = 1$ and for which $E(X)$ exists, then $P(X \geq t) \leq E(X)/t$

Expectation of Functions of Random Variables

Thoerem A Suppose that $Y = g(X)$. if X is discrete with frequency function $p(x)$ then $E(Y) = \Sigma_x g(x)p(x)$ if X is continuous with frequency function $f(x)$ then $E(Y) = \int_{-\infty}^{\infty} g(x)f(x)dx$

Thoerem B Suppose that $X_1, \dots, X_n$ are jointly distributed random variables and $Y = g(X_1, \dots, X_n)$. if X_i are discrete with frequency function $p(x_1, \dots, x_n)$ then $E(Y) = \Sigma_{x_1,\dots,x_n} g(x_1, \dots, x_n)p(x_1, \dots, x_n)$ if X_i are continuous with frequency function $f(x_1,\dots,x_n)$ then $E(Y) = \int\cdots\int g(x_1,\dots,x_n)f(x_1,\dots,x_n)dx_1\dots dx_n$

Corollary A: If X and Y are independent random variables and g and h are fixed functions, the $E[g(X)h(Y)] = \{E[g(X)]\}\{E[h(Y)]\}$, provided that the expectations on the right-hand side exists

Expectations of Linear Combinations of Random Variables:

Theorem A: if $X_1,\dots,X_n$ are jointly distributed random variables with $E(X_i)$ and Y is a linear funciton of the X_i. $Y = a + \Sigma_{i=1}^n b_i X_i$ then $E(Y) = a + \Sigma_{i=1}^n b_i E(X_i)$

Variance: if X is a random variable with expected value $E(X)$, the variance of X is $Var(X) = E\left([X - E(X)]^2\right)$. Provided that the expectation exists. The standard deviation of X is the square root of the variance.

if X_i are discrete with frequency function $p(x)$ with mean μ then $Var(Y) = \Sigma_i (x_i - \mu)^2 p(x_i)$ if X_i are continuous with frequency function $f(x)$ with mean μ then $Var(Y) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$

Theorem A: If $Var(X)$ exists, and $Y = a+bX$, then $Var(Y) = b^2 Var(X)$

Theorem B: If $Var(X)$ exists, can be calculated ad $Var(X) = E(X^2) - [E(X)]^2$

Theorem C: Let X is a random variable with mean μ and variance σ^2. Then for any $t > 0$: $P(|X-\mu| > t) \leq \frac{\sigma^2}{t^2}$

Corollary A: if $Var(X) = 0$, then $P(X = \mu) = 1$

Error Measurement: if an expetiment X is modeled as $X = x_0 + \beta + \epsilon$. we have $E(X) = x_0 + \beta$ and $Var(X) = \sigma^2$. x_0 is real value, β is constant error, ϵ is random error. The mean squared error $MSE = E\left[(X - x_0)^2\right]$ **Theorem A**: $MSE = \beta^2 + \sigma^2$

Covariance: The covariance of two random variables measure is a measure of the jion probability of 2 or more variables. if X and Y are jointly distributed random variables with expectectations μ_X and μ_Y, respectively. The covariance of X and Y is $Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = E(XY) - E(X)E(Y)$. If X and Y are independent, then $E(XY) = E(X)E(Y)$ and $Cov(X,Y) = 0$

Theorem A: Suppose that $U = a + \Sigma_{i=1}^n b_i X_i$ and $V = c + \Sigma_{j=1}^m d_j Y_j$ Then $Cov(U,V) = \Sigma_{i=1}^n \Sigma_{j=1}^m b_i d_j Cov(X_i, Y_j)$

$Var(X+Y) = Cov(X+Y, X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$

Corollary A: $Var(a + \Sigma_{i=1}^n b_i X_i) = \Sigma_{i=1}^n \Sigma_{j=1}^n b_i b_j Cov(X_i, X_j)$

Corollary B: $Var(\Sigma_{i=1}^n X_i) = \Sigma_{i=1}^n Var(X_i)$, if X_i are independent

Correlation coefficient: If X and Y are jointly distributed random variables with existingi variances and covariances that are non zero. Then the correlation of X and Y, denoted $\rho = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$ $\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$ and $\sigma_{XY} = \rho\sigma_X\sigma_Y$

Theorem B: if $-1 \leq \rho \leq 1$. Furthermore, $\rho = \pm 1$ if and only if $P(Y = a + bX) = 1$ for some constants a and b.

Conditional Expectation: Suppose Y and X are discrete random variables and that the conditional frequency function of Y given x is $p_{Y|X}(y|x)$. The conditional expectation of Y given $X = x$ is $E(Y|X=x) = \Sigma_y y p_{Y|X}(y|x))$ For the continuous case: $E(Y|X=x) = \int y f_{Y|X}(y|x))dy$. Or generally

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

Theorem A: $E(Y) = E[E(Y|X)]$

Theorem B: $Var(Y) = Var[E(Y|X)] + E[Var(Y|X)]$