Contents

Chapter 1

Distributions

1.1 Poisson

Expresses the probability of a given number of events occuring during a fixed interval of time or space if these events occur with a known constant rate independently of the time since the last event.
$\frac{\lambda^k e^{-\lambda}}{k!}$

1.2 Geometric

The probability distribution of the number X of bernoulli trials needed to get one success.
$(1-p)^{k-1}p$
$1-(1-p)^k$

1.3 Binomial

distribution of the number of successes in a sequence of n independent bernoulli trials.
$\binom{n}{k}p^k(1-p)^{n-k}$

1.4 Negative Binomial

number of successes in a sequence of iid bernoulli trials before a specified number of failures (r).
$\binom{k+r-1}{k}(1-p)^r p^k$

$$\text{Negative Binomial Mgf } \left(\frac{p}{1-(1-p)e^t}\right)^r$$

1.5 Hypergeometric

The result of each draw (the elements of the population being sampled) can be classified into two mutually exclusive categories ie pass/fail.
The probability of a success changes on each draw, as each draw decreases the population (sampling without replacement from a finite population)
N=population size
K=number of success in the population
n= number of draws
k=number of observed successes
$\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$

1.6 Exponential

describes the waiting time between Poisson events, Memoryless
$\lambda e^{-\lambda x}$
$1 - e^{-\lambda x}$

1.7 Normal

$\frac{1}{\sqrt{2\pi}\sigma}e^{-\dfrac{(x-\mu)^2}{2\sigma^2}}$ **Beta**
$\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$
where $B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

1.8 Uniform

Uniform Continuous symmetric probability distribution where all intervals of the same length are equally probable.
$\frac{1}{b-a}$
$\frac{x-a}{b-a}$

Uniform Discrete
a symmetric probability distribution where a finite number of values are equally likely to be observed. Every one of n values has an equal probability 1/n.
$\frac{1}{n}$

1.9 Gamma Distribution

Gamma Function: $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}e^{-t}\ dt$
$\Gamma(\alpha+1) = \alpha\Gamma(\alpha)\ \alpha > 0$
$\Gamma(n) = (n-1)! \quad n \in \mathbb{Z}$
$\Gamma(1/2) = \sqrt{\pi}$
$f(x|\alpha,\beta) = \dfrac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1}e^{-x/\beta}$
α is the shape parameter, influences the peakedness of the distribution
β is the scale parameter, influences the spread of the distribution
$EX^v = \dfrac{\beta^v\Gamma(v+\alpha)}{\Gamma(\alpha)}$
$\Gamma(\alpha+v) = \int_0^\infty x^{v+\alpha-1}e^{-x}\ \mathrm{d}x$
$EX = \alpha\beta$ alternatively $a/\lambda\ \lambda = 1/\beta$
$Var(X) = ab^2$ alternatively a/λ^2
$\int_0^\infty e^{-x^2/2}\ dz = \dfrac{\sqrt{2\pi}}{2} = \sqrt{\dfrac{\pi}{2}}$
$\int_0^\infty x^2e^{-x^2}$ is the same

1.10 Beta Distribution

sample space: $(0,1)$
$f(x|\alpha,\beta) = \dfrac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}$
Beta Function: $B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\ \mathrm{d}x$
$B(\alpha,\beta) = \dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$
$EX^n = \dfrac{B(\alpha+n,\beta)}{B(\alpha,\beta)} = \dfrac{\Gamma(\alpha+n)\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+n)\Gamma(\alpha)}$
$E(X) = \dfrac{a}{a+b}$
$Var(X) = \dfrac{ab}{(a+b)^2(a+b+1)}$
$beta(1,1) = U(0,1)$

1.11 Mgfs

Specific Characteristic Functions

	mgf	cf
Bernoulli(p)	$pe^t + q$	$pe^{it} + q$
Binomial(n,p)	$(pe^t + q)^n$	$(pe^{it} + q)^n$
Poisson(λ)	$e^{\lambda(e^{t-1})}$	$e^{\lambda(e^{it}-1)}$
Geometric(p)	$pe^t/(1-qe^t)$	$pe^{it}/(1-qe^{it})$
Negbin(n,p)	$\left[\frac{pe^t}{1-qe^t}\right]^n$	$\left[\frac{pe^{it}}{1-qe^{it}}\right]^n$
Uniform(a,b)	$\frac{e^{tb}-e^{ta}}{t(b-a)}$	$\frac{e^{itb}-e^{ita}}{it(b-a)}$
Normal(μ,σ^2)	$e^{t\mu+\frac{1}{2}\sigma^2t^2}$	$e^{it\mu-\frac{1}{2}\sigma^2t^2}$
Exponential(λ)	$\frac{\lambda}{\lambda-t} = (1-\frac{t}{\lambda})^{-1}$	$(1-\frac{it}{\lambda})^{-1}$
Gamma(a,λ)	$(1-\frac{t}{\lambda})^{-a}$	$(1-\frac{it}{\lambda})^{-a}$

1.12 Location and Scale Families

$$f_{\mu,\sigma}(x) = \frac{1}{\sigma}f\left(\frac{x-\mu}{\sigma}\right)$$

$\mu \in \mathbb{R} \quad \sigma > 0$ is a location scale family
If $\mu = 0$ scale family
If $\sigma = 1$ location family
Properties: Let $Z \sim f(z)$ and $X = \sigma Z + \mu$ Then
X has pdf $f_{\mu,\sigma}$
$E(X) = \sigma E(Z) + \mu \quad Var(X) = \sigma^2 Var(Z)$

Chapter 2

Notes 13

2.1 Conditional Probability

$$f(y|x) = \frac{f(x,y)}{f(x)}$$

2.2 Exponential Families

A family of pdfs or pmfs is called an exponential family if it can be expressed as

$$f(x|\theta) = h(x)c(\theta)\exp\left(\sum_{i=1}^{k} w_i(\theta)t_i(x)\right)$$

2.3 Multinomial Distribution

$p(s_1, s_2, \ldots, s_k) = \dfrac{n!}{s_1!s_2!\ldots s_k!} p_1^{s_1} p_2^{s_2} \cdots p_k^{s_k}$
where $\sum_{i=1}^{k} s_1 = n$ and $\sum_{i=1}^{k} p_i = 1$

Chapter 3

Notes 14

3.1 Convolution

If X and Y are independent continuous r.v.s with pdfs $f_X(x)$ and $f_Y(y)$, then the pdf of $Z = X + Y$ is:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(w) f_Y(z-w)\, dw$$

3.2 Sum of Two Independent Poissons

$X \sim Pois(\lambda_1),\ Y \sim Pois(\lambda_2)$
$U = X + Y\ V = Y$
$X = U - V\ Y = V$
Joint PMF of U and V is:
$f_{U,V}(u,v) = f_{X,Y}(u-v,v) = \dfrac{e^{-\lambda_1}\lambda_1^{u-v}}{(u-v)!}\dfrac{e^{-\lambda_2}\lambda_2^{v}}{v!}$
The distribution of $U = X + Y$ is the marginal:
$f_U(u) = \sum_{v=0}^{u} \dfrac{e^{-\lambda_1}\lambda_1^{u-v}}{(u-v)!}\dfrac{e^{-\lambda_2}\lambda_2^{v}}{v!}$
$= \dfrac{e^{-(\lambda_1+\lambda_2)}}{u!} \sum_{v=0}^{u} \binom{u}{v} \lambda_1^{u-v}\lambda_2^{v}$

Because of the binomial theorem

$= \dfrac{e^{-(\lambda_1+\lambda_2)}}{u!}(\lambda_1+\lambda_2)^u$

$U \sim Pois(\lambda_1 + \lambda_2)$

3.3 Jacobian

$J(u,v)$ is the Jacobian of the transformation $(x,y) \to (u,v)$ given by:

$$J(u,v) = \frac{\partial(x,y)}{\partial(u,v)} = \begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{bmatrix}$$

3.4 Functions of Independent Random Variables

Let X and Y be independent r.v.s
Let $g : \mathbb{R} \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$ be functions
Then the r.v.s $U = g(X)$ and $V = h(Y)$ are independent

3.5 Ratio of Two Independent Normals

Let $X \sim N(0,1)$ and $Y \sim N(0,1)$
The ratio X/Y has the Cauchy distribution
Let $U = X/Y$ and $V = Y$ Then $X = UV$ and $Y = V$ $J(u,v) = v$
$f_{X,Y}(x,y) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} = \frac{1}{2\pi} e^{-(x^2+y^2)/2}$
$f_{U,V}(uv,v) = \frac{1}{2\pi} e^{-[(uv)^2+v^2]/2} * |v| = \frac{|v|}{2\pi} e^{-(u^2+1)v^2/2}$
$f_U(u) = \int_{-\infty}^{\infty} f_{UV}(u,v)\ dv = 2\int_0^{\infty} \frac{v}{2\pi} e^{-(u^2+1)v^2/2}\ dv$
$= \frac{1}{\pi} \int_0^{\infty} e^{-(u^2+1)z}\ dz = \frac{1}{\pi(u^2+1)}$

3.6 Sum of Two Independent Random Variables

Suppose X and Y are independent, find distribution of $Z = X + Y$
In general: $F_Z(z) = P(X + Y \leq z) = P(\{(x,y) \text{ such that } x + y \leq z\})$
Approaches:

- Bivariate transformation method (continuous and discrete)

- Discrete convolution:
  $f_Z(z) = \sum_{x+y=z} f_X(x) f_Y(y) = \sum_x f_X(x) f_Y(z-x)$

- Continuous convolution

- Mgf/cf method (continuous and discrete)
  $\phi_Z \theta = \phi_X(\theta)\phi_Y(\theta)$
  $Z = X - Y \quad \phi_Z \theta = \phi_X(\theta)\phi_Y(-\theta)$

Chapter 4

Notes 15

4.1 Conditional Expectation and Variance

For two r.v.s X and Y with conditional pdf $f_{Y|X}(y|x)$ the conditional expectation of $g(Y)$ give $X = x$ is:
$h(x) = E[g(Y)|x] = \int_{-\infty}^{\infty} g(y) f_{Y|X}(y|x)\ dy$
$h(X) = E[g(Y)|X]$
Iterative Expectation Formula
$EX = E(E(X|Y))$
Variance
$Var[g(Y)] = E[g(y) - E(g(Y))]^2$

$VarX = E(Var(X|Y)) + Var(E(X|Y))$
$Var(g(Y)|X) = E\{[g(Y) - E(g(Y)|X)]^2|X\}$
where both expectations are taken with respect to $f_{Y||X}(y)$

- $E(Var(X|Y)) = E\{[X - E(X|Y)]^2\}$

- $Var(E(X|Y)) = E\{[E(X|Y) - EX]^2\}$

4.2 Covariance and Correlation

$Cov(X,Y) = E[(X - \mu_X)(Y - \mu_Y)] = \sigma_{XY}$
Correlation $= \rho_{XY} = \dfrac{Cov(X,Y)}{\sqrt{VarX\ VarY}} = \dfrac{\sigma_{XY}}{\sigma_X \sigma_Y}$
$= E\left[\left(\dfrac{X - \mu_X}{\sigma_X}\right)\left(\dfrac{Y - \mu_Y}{\sigma_Y}\right)\right]$

X and Y are uncorrelated iff:
$Cov(X,Y) = 0$ or equivalently $\rho_{XY} = 0$

$Cov(X,Y) = E(XY) - E(X)E(Y)$
If X and Y are independent and $Cov(X,Y)$ exists, then $Cov(X,Y) = 0$
If X and Y are uncorrelated this does not imply independence.

4.3 Linear Combinations

$Cov(aX + B_Y, Z) = aCov(X,Z) + bCov(Y,Z)$

$$Var(aX + bY) = a^2Var(X) + b^2Var(Y) + 2abCov(X,Y)$$

$$Corr(aX + b, cY + d) = \frac{ac}{|ac|}Corr(X,Y)$$

4.4 Standard Bivariate Normal

$f_{XY}(x,y) = \dfrac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\dfrac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right]$
Both X and Y have marginal distributions are $N(0,1)$
Correlation of X and Y is ρ
Conditional Distribution are normal:
$Y|X \sim N(\rho X, 1-\rho^2) \quad X|Y \sim N(\rho Y, 1-\rho^2)$
The means are the regression lines of Y on X and X on Y respectively.

4.5 Bivariate Normal

Let $\tilde{X}$ and $\tilde{Y}$ have a standard bivariate normal distribution with correlation ρ
Let $X = \mu_X + \sigma_X\tilde{X} \quad Y = \mu_Y + \sigma_Y\tilde{Y}$
Then (X,Y) has the bivariate normal density:
$f_{XY}(x,y) = \left(2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}\right)^{-1} \exp\left\{-\dfrac{1}{2(1-\rho^2)}\left[\left(\dfrac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\dfrac{x-\mu_X}{\sigma_X}\right)\left(\dfrac{y-\mu_Y}{\sigma_Y}\right) + \left(\dfrac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\}$

Marginal distributions: $N(\mu_X, \sigma_X^2) \quad N(\mu_Y, \sigma_Y^2)$
$Corr(X,Y) = \rho$

Conditional distributions are normal:
$Y|X \sim N[\mu_Y + \rho(\sigma_Y/\sigma_X)(x-\mu_X), \sigma_Y^2(1-\rho^2)]$

Distribution of $aX + bY$ is:
$N(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\rho\sigma_X\sigma_Y)$

4.6 Multivariate Distributions

$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$
If $\boldsymbol{X}$ is discrete then:
$P(\boldsymbol{X} \in A) = \sum_{\boldsymbol{X} \in A} f(\boldsymbol{X})$

where $f(\boldsymbol{X})$ is the joint pmf
If $\boldsymbol{X}$ is continuous then:
$P(\boldsymbol{X} \in A) = \int \cdots \int_A f(x_1, \dots, x_n) dx_1, \dots dx_n$

4.7 Marginals and Conditionals

The **marginal** pdf or pmf of any subset of coordinates is found by integrating or summing the joint pdf or pmf over all possible values of the other coordinates.

The **conditional** pdf or pmf of a subset of coordinates given the values of the remaining coordinates is found by dividing the full joint pdf or pmf by the joint pdf or pmf of the remaining variables.

4.8 Multivariate Independence

Independent Random Vectors:
Let $\boldsymbol{X_1}, \dots, \boldsymbol{X_n}$ be random vectors with joint pdf or pmf $f(\boldsymbol{X_1}, \dots, \boldsymbol{X_n})$
Let $f\boldsymbol{X_j}(\boldsymbol{x_j})$ be the marginal pdf or pmf of $\boldsymbol{X_j}$.
Then $\boldsymbol{X_1}, \dots, \boldsymbol{X_n}$ are **mutually independent** random vectors if:
$\forall\ (\boldsymbol{X_1}, \dots, \boldsymbol{X_n})$: $\quad f(\boldsymbol{X_1}, \dots, \boldsymbol{X_n}) = \prod_{j=1}^n f\boldsymbol{X_j}(\boldsymbol{x_j})$

4.9 Multinomial

Let n and m be positive integers and let $p_1, \dots, p_n$ be probabilities summing to one. Then the random vector $(X_1, \dots, X_n)$ has a multinomial distribution with m trials and cell probabilities $p_1, \dots, p_n$ if its joint pmf is:

$$\begin{aligned} f(x_1, \dots, x_n) &= \binom{m}{x_1, \cdots, x_n} p_1^{x_1} \cdots p_n^{x_n} \\ &= \frac{m!}{x_1! \dots x_n!} p_1^{x_1} \cdots p_n^{x_n} \\ &= m! \prod_{j=1}^n \frac{p_j^{x_j}}{x_j!} \end{aligned}$$

for $x_1 = 0, \dots, m \quad i = 1, \dots, n \quad x_1 + \cdots + x_n = m$

Chapter 5

Notes 16

5.1 Inequalities

Chebychev Inequality
$P[g(X) \geq r] \leq \frac{E[g(X)]}{r}$
If X is nonnegative and g is a positive non-decreasing function then:
$P\{X \geq a\} \leq \frac{E[g(X)]}{g(a)}$
Special Cases:
$X \geq 0 \quad P\{X \geq a\} \leq \frac{E(e^{tX})}{e^{ta}}$
L^p Space- consists of all r.v.s whose p^{th} absolute power is integrable, $E(|X|^p) < \infty$
Triangle Inequality
$|a + b| \leq |a| + |b|$
Convex Functions
A function $g : I \to R$ is convex for any $\lambda \in [0, 1]$ and any points x and y in I
$g[\lambda x + (1 - \lambda)y] \leq \lambda g(x) + (1 - \lambda)g(y)$
A differentiable function g is convex iff it lays above all tangents.
A twice differentiable function g is convex iff its second derivative is non-negative
concave if -g is convex on I
Jensen's Inequality
Let $X \in L^1$ and $g(x)$ be a convex function where $E[g(X)]$ exists. Then:
$E[g(X)] \geq g[EX]$
with equality iff for every line $a + bx$ tangent to $g(x)$ at $x = EX, P[g(X) = a + bX] = 1$
direction of inequality is reversed if g is concave
Young's Inequality
Let $a, b > 0$ and $p, q > 1$ with $1/p + 1/q = 1$ Then:
$\frac{a^p}{p} + \frac{b^q}{q} \geq ab$

with equality iff $a^p = b^q$
Holder's Inequality
Suppose $X \in L^p, Y \in L^q$ where $p, q > 1$ and $1/p + 1/q = 1$ Then:
$E[|XY|] \leq [E|X|^p]^{1/p} E[|Y|^q]^{1/q}$
with equality if $X^p = cY^q$ for some $c \in \mathbb{R}$
Cauchy-Schwartz Inequality
corollary of Holders where $p = q = 2$
$E[|XY|] \leq [E|X|^2]^{1/2} E[|Y|^2]^{1/2} = \sqrt{E[X^2]E[Y^2]}$
with equality if $X = cY$
Lyapunov's Inequality
corrallary of Holders
for $1 \leq r \leq s$ and $X \in L^s$
$E[|X|^r]^{1/r} \leq E[|X|^s]^{1/s}$
Application of Cauchy-Schwartz
$p = \dfrac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$
Then $|p| \leq 1$ with equality iff $Y - \mu_Y = c(X - \mu_X)$
Minkowski's Inequality
Suppose $X, Y \in L^p, p \geq 1$ Then $(X + Y) \in L^p$ and
$[E|X + Y|^p]^{1/p} \leq [E|X|^p]^{1/p} + [E|Y|^p]^{1/p}$

5.2 Order Statistics

Distribution of the Maximum
The cdf of $Z = max(Y_1, \ldots, Y_n)$ is

$$\begin{aligned}
F_Z(z) &= P\{Z \leq z\} \\
&= P\{Y_1 \leq z, Y_2 \leq z, \ldots, Y_n \leq z\} \\
&= \prod_{j=1}^{n} P\{Y_j \leq z\} \text{ indep} \\
&= F_Y(z)^n \text{ ident. distrib}
\end{aligned}$$

Thus the pmf is:
$f_Z(z) = nF_Y(z)^{n-1} f_Y(z)$
Distribution of the Minimum
$W = min(Y_1, \ldots, Y_n)$
$F_W(w) = 1 - (1 - F_Y(w))^n$
$f_W(w) = n(1 - F_Y(w))^{n-1} f_Y(w)$
Order Statistics
Let $Y_1, Y_2, \ldots, Y_n$ be iid with pdf $f_Y(x)$
Order the observations:
$Y_{(1)} \leq Y_{(2)} \leq \cdots \leq Y_{(n)}$

The $Y_{(i)}$ are called order statistics. Minimum is $Y_{(1)}$ max is $Y_{(n)}$
We are interested in finding the distribution of an arbitrary $Y_{(i)}$ as well as the joint distributions of sets of $Y_{(i)}$s and $Y_{(j)}$s
ex: Range=$Y_{(n)} - Y_{(1)}$
r^{th} **order statistic**
We need to find the density of $Y_{(r)}$ at a value y
Consider 3 intervals $(-\infty, y), [y, y+dy), [y+dy, \infty)$
The number of observations in each of the intervals follows the tri-nomial distribution:
$f(s_1, s_2, s_3) = \dfrac{n!}{s_1!s_2!s_3!} p_1^{s_1} p_2^{s_2} p_3^{s_3}$
The event that $y \leq Y_{(r)} < y + dy$ is the event we have:
$(r-1)$ observations are less than y,
$(n-r)$ observations are greater than y
1 observation is in interval $y, y+dy$
In the trinomial distribution this corresponds to:
$s_1 = r-1, s_2 = 1, s_3 = n-r$
$p_1 = F_Y(y), p_2 = f_Y(y)dy, p_3 = 1 - F_Y(y+dy)$
Theorem 5.4.6
Let $X_{(1)}, \ldots, X_{(n)}$ denote the order statistics of a random sample, $X_1, \ldots, X_n$ from a continuous population with cdf $F_X(x)$ and pdf $f_X(x)$. Then the joint pdf of $X_{(i)}$ and $X_{(j)}$, $1 \leq i < j \leq n$, is:

$$f_{X_{(i)},X_{(j)}}(u,v) = \frac{n!}{(i-1)!(j-1-i)!(n-j)!} f_X(u) f_X(v) [F_X(u)]^{i-1} [F_X(v) - F_X(u)]^{j-1-i} [1 - F_X(v)]^{n-j}$$

for $-\infty < u, v < \infty$
The joint pdf of all the order statistics is:

$$f_{X_{(1)} \ldots, X_{(n)}}(x_1, \ldots, x_n) = \begin{cases} n! f_X(x_1) \cdots f_X(x_n) & -\infty < x_1 < \cdots < x_n < \infty \\ 0 \quad \text{otherwise} \end{cases}$$