**Probability- the good parts version**

## I. Random variables and their distributions; continuous random variables.

A random variable (r.v) $X$ is continuous if its distribution is given by a probability density function (pdf) $f(x)$ that is positive on an interval. For real numbers $a < b$,

$$P(a < X < b) = \int_a^b f(x)dx.$$

Random variables $X$ and $Y$ are jointly continuous if there's a joint density function $f(x, y)$ such that

$$P(a < X < b, c < Y < d) = \int_a^b \int_c^d f(x, y)dydx.$$

The marginal density for $X$ is found by integrating out the $y$ and vice-versa for $Y$. $X$ and $Y$ are called independent if their joint density is the product of their marginals. In this case, it follows that $P(a < X < b, c < Y < d) = P(a < X < b)P(c < Y < d)$.

The cumulative distribution function (cdf) of $X$ is the function $F$,

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t)dt.$$

Given a function $g()$, the expected value of $g(X) =$

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

In particular, the mean of $X= E(X)= \mu$; the variance of $X= V(X)= \sigma^2= E(X-\mu)^2= EX^2-\mu^2$; and the standard deviation $\sigma = \sqrt{\sigma^2}$.

The moment generating function of $X$ is

$$M_X(t) = E(e^{Xt}) = \int_{-\infty}^{\infty} e^{xt} f(x)dx.$$

Properties of the moment generating function:

(1) $M_X^{(n)}(0) = \frac{d^n}{dt^n} M_X(t),$ evaluated at t=0$= EX^n$.

(2) If $X_1, X_2, ..., X_n$ are independent random variables then

$$M_{X_1+X_2+...+X_n}(t) = M_{X_1}(t) * M_{X_2}(t) * \cdots * M_{X_n}(t).$$

(3) The m.g.f. specifies the distribution: if $M_X(t) = M_Y(t)$, then $X$ and $Y$ have the same distribution.

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

The standard class of continuous distributions.

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

density: $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}, \ -\infty < x < \infty.$

mean: $E(X) = \mu.$

variance: $V(X) = \sigma^2$.

mgf: $M(t) = e^{\mu t + \sigma^2 t^2/2}$.

If $X \sim N(\mu, \sigma^2)$, then $(X - \mu)/\sigma \sim N(0, 1)$.

If $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$ are independent, then $aX + bY \sim N(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2)$.

(2) $X \sim$ Chi-Square with n degrees of freedom ($X \sim \chi^2(n)$) if $X \sim$ Gamma$(n/2, 1/2)$.

mean: $E(X) = n$.

variance: $V(X) = 2n$.

mgf: $M(t) = (\frac{1}{1-2t})^{n/2}, t < 1/2$.

(3) $X \sim$ Exponential$(\lambda)$.

density: $f(x) = \lambda e^{-\lambda x}, x \geq 0$.

mean: $E(X) = 1/\lambda$.

variance: $V(X) = 1/\lambda^2$.

mgf: $M(t) = (\frac{\lambda}{\lambda - t}), t < \lambda$.

$P(X > x) = e^{-\lambda x}$, for $x > 0$.

Note: There are 2 common conventions followed for the Exponential($\lambda$): (i) $\lambda$ is the parameter in the density as above and the mean is $1/\lambda$, and (ii) $\lambda$ is the mean and the density is $f(x) = (1/\lambda)e^{-x/\lambda}$; the interpretation in a particular problem should be clear from context.

(4) $X \sim \text{Gamma}(\alpha, \lambda)$.

density: $f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha - 1} e^{-\lambda x}, x \geq 0$.

mean: $E(X) = \alpha/\lambda$.

variance: $V(X) = \alpha/\lambda^2$.

mgf: $M(t) = (\frac{\lambda}{\lambda - t})^\alpha, t < \lambda$.

$Note : \Gamma(\alpha) = \int_0^\infty t^{\alpha - 1} e^{-t} dt$. $\Gamma(n) = (n-1)!$ when $n$ is a positive integer.

When $X_1, X_2, \ldots, X_n \sim \text{Exponential}(\lambda)$ are independent, $X_1 + X_2 + \cdots + X_n \sim \text{Gamma}(n, \lambda)$. In particular, Exponential($\lambda$) = Gamma($1, \lambda$).

(5) $X \sim U[a, b]$

density: $f(x) = 1/(b-a)$ for $a < x < b$.

mean: $E(X) = (a+b)/2$.

variance: $V(X) = (b-a)^2/12$.

mgf: $M(t) = \frac{e^{bt} - e^{at}}{(b-a)t}, -\infty < t < \infty$.

If [c,d] $\subset$ [a,b], then $P(c \leq X \leq d)$ is Length[c,d]/Length[a,b].

## II. Random variables and their distributions; discrete random variables.

A discrete rv $X$ takes on a finite or countable number of values. Probabilities are computed using a frequency function $p(k) = P(X = k)$; this is also called a probability density function (pdf) or probability mass function.

$$P(a < X < b) = \sum_{k \in (a,b)} p(k).$$

Given a function $g$, the expected value of $g(X) =$

$$E(g(X)) = \sum_k g(k)p(k);$$

in particular, the moment generating function of $X$ is

$$M_X(t) = E(e^{Xt}) = \sum_k e^{kt}p(k).$$

The standard class of discrete distributions.

(1) $X \sim \text{Bernoulli}(p)$

frequency function: $p(1) = p, p(0) = q = 1 - p$.

mean: $E(X) = p$.

variance: $V(X) = pq$.

mgf: $M(t) = (q + pe^t), -\infty < t < \infty$.

(2) $X \sim \text{Binomial}(n, p)$

frequency function: $p(k) = \binom{n}{k} p^k q^{n-k}, k = 0, 1, \ldots, n$.

mean: $E(X) = np$.

variance: $V(X) = npq$.

mgf: $M(t) = (q + pe^t)^n, -\infty < t < \infty$.

If $X$ is the number of successes in $n$ independent Bernoulli trials, then $X \sim \text{Binomial}(n, p)$.

(3) $X \sim \text{Geometric}(p)$. There are two definitions for the Geometric($p$) distribution. (I) $X$ is the number of failures required to see the first success in a sequence of Bernoulli trials and (II) $X$ is the number of trials required to see the first success in a sequence of Bernoulli trials. If X and Y represent have these respective distributions, then $Y = X + 1$. We give results separately for the two definitions.

Case I.

frequency function: $p(k) = pq^k, k = 0, 1, \ldots$, where $q = 1 - p$.

mean: $E(X) = q/p$.

variance: $V(X) = q/p^2$.

mgf: $M(t) = \frac{p}{1-qe^t}, -\infty < t < \ln(1/q)$.

Case II.

frequency function: $p(k) = pq^{k-1}, k = 1, 2, \ldots$, where $q = 1 - p$.

mean: $E(X) = 1/p$.

variance: $V(X) = q/p^2$.

mgf: $M(t) = \frac{pe^t}{1-qe^t}, -\infty < t < \ln(1/q)$.

(4) $X \sim$ Negative Binomial$(r, p)$. Again, there are two definitions for the Negative Binomial$(r, p)$ distribution. (I) $X$ is the number of failures before the rth success in a sequence of Bernoulli trials and (II) $Y$ is the number of trials required to see the rth success in a sequence of Bernoulli trials. If X and Y have these respective distributions, then $Y = X + r$. We give results separately for the two definitions.

Case I.

frequency function: $p_X(k) = \binom{k+r-1}{r-1} p^r q^k, k = 0, 1, \ldots$, where $q = 1 - p$.

mean: $E(X) = rq/p$.

variance: $V(X) = rq/p^2$.

mgf: $M(t) = (\frac{p}{1-qe^t})^r, -\infty < t < \ln(1/q)$.

Case II.

frequency function: $p_Y(k) = \binom{k-1}{r-1} p^r q^{k-r}, k = r, r+1, \ldots$, where $q = 1-p$.

mean: $E(X) = r/p$.

variance: $V(X) = rq/p^2$.

mgf: $M(t) = (\frac{pe^t}{1-qe^t})^r, -\infty < t < \ln(1/q)$.

(5) $X \sim \text{Poisson}(\lambda)$

frequency function: $p(k) = \frac{e^{-\lambda}\lambda^k}{k!}, k = 0, 1, \ldots$.

mean: $E(X) = \lambda$.

variance: $V(X) = \lambda$.

mgf: $M(t) = e^{\lambda(e^t-1)}, -\infty < t < \infty$.

## III. General Properties.

$E(X)$

$E(aX + bY) = aE(X) + bE(Y)$ for any random variables $X$ and $Y$.

$E(XY) = E(X)E(Y)$ if $X$ and $Y$ are independent.

$V(X)$ and $Cov(X, Y)$

$V(X) = E[(X - \mu)^2] = E(X^2) - \mu^2$.

$V(aX + b) = a^2V(X)$.

$V(X + Y) = V(X) + V(Y) + 2Cov(X, Y)$ for any random variables $X$ and $Y$.

$V(X + Y) = V(X) + V(Y)$ if $X$ and $Y$ are independent.

$Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - E(X)E(Y)$.

$Cov(X, Y) = 0$ if $X$ and $Y$ are independent.

$Cov[X, X) = V(X)$.

$Cov(aX, bY) = abCov(X, Y)$.

$Cov(X + Y, U + V) = Cov(X, U) + Cov(X, V) + Cov(Y, U) + Cov(Y, V)$.

## Sampling

If $\{X_i\}$ are $n$ independent, identically distributed random variables with $E(X_i) = \mu$ and $V(X_i) = \sigma^2$ and $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is the sample mean, then:

$E(\bar{X}) = \mu$ and $V(\bar{X}) = \sigma^2/n$.

## Central limit theorem

If $\{X_i\}$ are $n$ independent, identically distributed random variables with $E(X_i) = \mu$ and $V(X_i) = \sigma^2$, then

$\sum_{i=1}^{n} X_i$ approximately $\sim N(n\mu, n\sigma^2)$ as $n \to \infty$.

$\bar{X}$ approximately $\sim N(\mu, \sigma^2/n)$ as $n \to \infty$.

## Joint and conditional distributions

If $X$ and $Y$ have joint pdf $f_{X,Y}(x,y)$, then

$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$ or $\sum_{\text{all y}} f_{X,Y}(x,y)$;

$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$ or $\sum_{\text{all x}} f_{X,Y}(x,y)$;

$f_{X|Y=y}(x) = f_{X,Y}(x,y)/f_Y(y)$;

$f_{Y|X=x}(y) = f_{X,Y}(x,y)/f_X(x)$;

$f_{X,Y}(x,y) = f_X(x)f_Y(y)$ if and only if $X$ and $Y$ are independent.

$$X_{\max},\ X_{\min}$$

If $\{X_i\}$ are $n$ independent, identically distributed random variables with pdf $f_X(x)$, then

$f_{X_{\max}}(x) = nf_X(x)[F_X(x)]^{n-1}$

$f_{X_{\min}}(x) = nf_X(x)[1 - F_X(x)]^{n-1}$