Appendix B
Fundamentals of Probability

Jim Bang

Probability Distributions

Probability Mass Functions of Discrete Random Variables

Probability mass function, $Pr(x)$, maps all possible values of X into the probabilities that $X = x$.

Table 1: Discrete Probability Distributions in R: x is the value of the variable; p is the cumulative probability; and R is the number of random draws. Other arguments represent parameters of the distribution.

Name	PMF [$Pr(x = x)$]	CDF [$P(x <= x)$]	Quantile [$P^{-1}(p)$]	Random #s
Bernoulli	dbinom(x, 1, π_s)	pbinom(x, 1, π_s)	qbinom(x, 1, π_s)	rbinom(R, 1, π_s)
Binomial	dbinom(x, n, π_s)	pbinom(x, n, π_s)	qbinom(p, n, π_s)	rbinom(R, n, π_s)
Hypergeometric	dhyper(x, S, F, n)	phyper(x, S, F, n)	qhyper(p, S, F, n)	rhyper(R, S, F, n)
Poisson	dpois(x, λ)	dpois(x, λ)	qpois(p, λ)	rpois(R, λ)
Geometric	dgeom(x, π_s)	dgeom(x, π_s)	qgeom(p, π_s)	rgeom(R, π_s)

Probability Density Functions of Continuous Random Variables

Probability density function, $f(x)$, measures how tightly packed the probability is around x, *not* the probability at that point ($Pr[X = x] = 0$).

Table 2: Coninuous Probability Distributions in R: x is the value of the variable; p is the cumulative probability; and R is the number of random draws. Other arguments represent parameters of the distribution.

Name	PMF $[f(x)]$	CDF $[P(x <= x)]$	Quantile $[P^{-1}(p)]$	Random #s
Uniform	dunif(x, min, max)	punif(x, min, max)	qunif(p, min, max)	runif(R, min, max)
Logistic	dlogis(x, μ)	plogis(x, μ)	qlogis(p, μ)	rlogis(R, μ)
Exponential	dexp(x, λ)	pexp(x, λ)	qexp(p, λ)	rexp(R, λ)
Std. Normal	dnorm(x, 0, 1)	pnorm(x, 0, 1)	qnorm(p, 0, 1)	rnorm(R, 0, 1)
Normal	dnorm(x, μ, σ)	pnorm(x, μ, σ)	qnorm(p, μ, σ)	rnorm(R, μ, σ)
Lognormal	dlnorm(x, μ, σ)	plnorm(x, μ, σ)	qlnorm(p, μ, σ)	rlnorm(R, μ, σ)
χ^2	dchisq(x, n)	pchisq(x, n)	qchisq(p, n)	rchisq(R, n)
Student's t	dt(x, df)	dt(x, df)	qt (p, df)	rt(R, df)
F	df(x, df_1, df_2)	pf(x, df_1, df_2)	qf(p, df_1, df_2)	rf(R, df_1, df_2)

Cumulative Probability Distribution of a Random Variable

Cumulative probabilities for a discrete R.V.

$$P(X \leq x^*) = \sum_{x=-\infty}^{x^*} p(x)$$

Cumulative probabilities for a continuous R.V.

$$P(X \leq x^*) = F(x^*) = \int_{-\infty}^{x^*} f(x)dx$$

= The area under f to the left of x.

$$P(X = x^*) = \int_{x^*}^{x^*} f(x)dx = F(x^*) - F(x^*) = 0$$

Expected Value

The first moment of X is $E[X]$, or μ.

Discrete Random Variable

$$E(X) = \sum_{x=-\infty}^{\infty} xp(x) = \vec{x} \cdot \vec{p}(x)$$

For a random variable, y, defined on a uniform *discrete* distribution, py, from zero to ten, calculate $E(y)$.

```
y <- c(0:10)
py <- rep(1/length(y), length(y))
y%*%py
```

```
##      [,1]
## [1,]    5
```

Continuous Random Variable

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

For a random variable, x, defined on a uniform *continuous* distribution between zero and ten, calculate $E(x)$.

```
integrate(function(x) x/10, 0, 10)
```

```
## 5 with absolute error < 5.6e-14
```

Notice that $E(X^2) > E(X)^2$. This because x^2 is a convex function, and this property is known as Jensen's Inequality.

Expected Value of a Function of a Random Variable

Linear functions and expectation

1. Constants can factor out of the expectation: $E(c) = c$ for any constant, c.
2. Addition is separable inside expectation: $E(X + Y) = E(X) + E(Y)$.
3. For any *linear* function, f, $E[f(X)] = f[E(X)]$, i.e. $E(a + bX) = a + bE(X)$ for any constants, a and b.
4. For any *convex* function, f, $E[f(X)] > f[E(X)]$, e.g. $E(X^2) > E(X)^2$; For any *concave* function, f, $E[f(X)] < f[E(X)]$, e.g. $E[ln(X)] < ln[E(X)]$.

For a random variable, x, defined on a uniform *continuous* distribution between zero and ten, calculate $E(x)^2$ and $E(x^2)$.

```
# Calculate E(x), E(X)^2, and E(X^2)
integrate(function(x) x/10, 0, 10)$value^2
```

```
## [1] 25
```

```r
integrate(function(x) (x^2)/10, 0, 10)$value
```

```
## [1] 33.33333
```

Mean, Median, and Mode

An important property of the mean is that the estimator for it represents the parameter that minimizes the squared differences of the observed data from it.

$$\hat{\mu} = \underset{\theta}{\operatorname{argmin}}\left\{\sum_{i=1}^{n}(x_i - \theta)^2\right\}$$

We can rewrite this property as the minimum of the squared absolute values (it doesn't change anything because the squared deviations are already positive).

$$\hat{\mu} = \underset{\theta}{\operatorname{argmin}}\left\{\sum_{i=1}^{n}|x_i - \theta|^2\right\}$$

We learn the *median* as the value of x that "splits" the distribution (or population or sample) in half. An interesting property of the median is that it minimizes the *absolute* differences (to the first power) of the observed data from itself.

$$\tilde{\mu} = \underset{\theta}{\operatorname{argmin}}\left\{\sum_{i=1}^{n}|x_i - \theta|\right\}$$

The *mode* measures the most frequent or most likely value of x. An interesting property of the mode is that it minimizes the *absolute* differences of the observed data from itself *raised to the zero power*.

$$\mathring{\mu} = \underset{\theta}{\operatorname{argmin}}\left\{\sum_{i=1}^{n}|x_i - \theta|^0\right\}$$

Notice that the distance function, $|x_i - \theta|^0$ equals 0 if $x_i = \theta$ and 1 if $x_i \neq \theta$, so assigning θ equal to the most frequent value minimizes the sum of these distances by setting the most values to zero instead of one.

Summarizing,

1. Mode: $\mathring{\mu} = \underset{\theta}{\operatorname{argmin}}\left\{\sum_{i=1}^{n}|x_i - \theta|^0\right\}$

2. Median: $\tilde{\mu} = \underset{\theta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} |x_i - \theta|^1 \right\}$
3. Mean: $\hat{\mu} = \underset{\theta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} |x_i - \theta|^2 \right\}$

Variance and Standard Deviation

The second moment of X is $E[X^2]$; the second *central* moment of X around its mean is $E[(X\breve{\ }\mu)^2]$, also known as $V(X)$ or σ_x^2 (the first central moment, $E[X-\mu]$ is always zero).

The standard deviation is the square root of the variance, or σ_x.

Discrete Random Variable

For a discrete uniform random variable from zero to ten, Z, calculate $Var(Z)$.

```
y <- c(0:10)
py <- rep(1/length(y), length(y))
(y - c(y%*%py))^2%*%py
```

```
##      [,1]
## [1,]   10
```

Continuous Random Variable

For a *continuous* uniform random variable from zero to ten, X, calculate $Var(Z)$.

```
mu <- integrate(function(x) x/10, 0, 10)$value
integrate(function(x) (x - mu)^2/10, 0, 10)
```

```
## 8.333333 with absolute error < 9.3e-14
```

Variance and Standard Deviation of a Linear Function of a Random Variable

Linear functions and variance.

1. Constants have zero variance: $V(c) = 0$ for any constant, c.
2. Addition distributes within the variance: $V(X+Y) = V(X) + V(Y) + 2COV(XY)$.
3. Coefficients factor out as squared values: $V(cX) = c^2V(X)$.
4. For any *linear* function, f, $V(a+bX) = b^2V(X)$ for any constants, a and b.

In the case of standard deviation, taking square roots gives: (1) $\sigma_c = 0$; (2) $\sigma_{(x+y)} = \sqrt{\sigma_x^2 + \sigma_y^2 + 2\sigma_{sy}}$; 3. $\sigma_{cX} = c\sigma_x$; and 4. $\sigma_{(a+bX)} = b\sigma_x$.

Other Moments of a Random Variable

Skewness

- 3rd moment: $E[X^3]$
- 3rd *central* moment: $E[(X \breve{\ } \mu)^3]$
- 3rd *standardized* moment: $E[(\frac{X \breve{\ } \mu}{\sigma})^3]$

Kurtosis

- 4th moment: $E[X^4]$
- 4th *central* moment: $E[(X \breve{\ } \mu)^4]$
- 4th *standardized* moment: $E[(\frac{X \breve{\ } \mu}{\sigma})^4]$