Lecture 2: Probability and Statistics (continued)

©Christopher S. Bretherton

Winter 2015

2.1 Expectation and moments

Expectation of a function $g(X)$ of a RV X is

$$E[g(X)] = \sum_{x:p(x)>0} g(x)p(x)dx \qquad \text{discrete RV } X$$

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx \qquad \text{continuous RV } X$$

The expectation of X is also called its **mean** μ_X, sometimes denoted $\overline{X}$.

Variance $\text{var}[X] = E[(X-\mu_X)^2 = E[X^2] - (E[X])^2$, whose square root is the **standard deviation** σ_X, a measure of the spread of X about its mean.

***n*'th moment** $E[X^n]$. The third moment is a measure of **skewness** or asymmetry of the PDF of X about its mean.

2.2 Examples of random variables

Bernoulli $P(X=1) = p$; $P(X=0) = q = 1-p$. $\mu_X = p$ and $\sigma_X = (pq)^{1/2}$. The sum of $N \geq 1$ independent identically-distributed Bernoulli random variables is a **binomial** distribution with parameters N and p.

Uniform distribution on (α, β):

$$f(x) = \frac{1}{\beta - \alpha}, \qquad \alpha < x < \beta.$$

$\mu_X = (\alpha+\beta)/2$ and $\sigma_X = (\beta-\alpha)/\sqrt{12}$; note how they scale with α, β.

Gaussian (or **normal**) distribution $n(\mu, \sigma)$, with mean μ and standard deviation σ:

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

$$F(a) = \int_{-\infty}^{a} f(x)dx = 0.5\left(1 + \text{erf}\left[\frac{a-\mu}{2^{1/2}\sigma}\right]\right)$$

Lognormal distribution on $0 < x < \infty$ with log-mean μ and log standard deviation σ:

$$\log(X) = n(\mu, \sigma), \qquad \mu_X = \exp\left(\mu + \frac{\sigma^2}{2}\right),\ \sigma_X = \mu_X\sqrt{\exp(\sigma^2) - 1}.$$

2.2.1 Generating random variables in Matlab

rand(m,n) returns an $m \times n$ matrix of random numbers from a uniform distribution on (0, 1).

randn(m,n) returns an $m \times n$ matrix of normally-distributed random numbers with mean 0 and standard deviation 1. Fig. 1 shows a histogram of the results of **randn(1,1000)**.

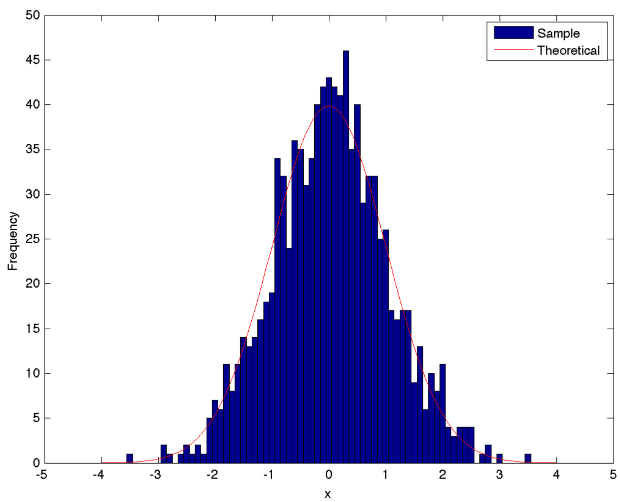

Figure 1: Histogram of 1000 samples of a normal distribution

random(name,params,[m,n,...)] (Statistics toolbox) returns an $m \times n \times ...$ array of random numbers with a pdf described by **name** and **params**, (e. g. 'Binomial',N,p or 'Lognormal',mu,sigma)

2.3 Joint distributions

Joint cumulative distribution of two RVs X and Y can be phrased in terms of their joint CDF

$$F(a, b) = P(X \leq a, Y \leq b)$$

Joint PDF $f(x, y)$ of two continuous RVs: $f(x, y)dxdy$ is the probability that $x - dx/2 < X < x + dx/2, y - dy/2 < Y < y + dy/2$.

Two RVs are independent iff

$$F(a, b) = F_X(a)F_Y(b)\ \forall\ a, b \qquad \text{or} \qquad f(x, y) = f_X(x)f_Y(y)\ \forall\ x, y$$

Covariance of X and Y:

$$\mathrm{cov}[X, Y] = E[(X - \overline{X})(Y - \overline{Y})]. \tag{2.3.1}$$

If X and Y are independent, $\mathrm{cov}[X, Y] = 0$ (but not necessarily vice-versa). Note $\mathrm{cov}[X, X] = \mathrm{var}[X]$ and $\mathrm{cov}[X, Y + Z] = \mathrm{cov}[X, Y] + \mathrm{cov}[X, Z]$.

Correlation coefficient

$$\mathrm{R}_{XY} = \frac{\mathrm{cov}[X, Y]}{\sigma_X \sigma_Y} \tag{2.3.2}$$

R lies between -1 and 1; $R = 1$ if $Y = X$ (perfect correlation), $R = -1$ if $Y = -X$ (perfect anticorrelation), and $R = 0$ if X and Y are independent. Unlike covariance, R is *not* additive.

The correlation coefficient is useful for describing how strongly X and Y are linearly related, but will not perfectly capture non-linear relationships between X and Y. In particular, unless X and Y are Gaussian, they can be uncorrelated ($R = 0$) yet still be dependent. For instance, let Θ be a uniformly distributed RV over $[0, 2\pi)$ and let $X = \cos(\Theta), Y = \sin(\Theta)$ (Fig. 2). Then X and Y each have mean zero and they are easily shown to be uncorrelated. However, for any given value x of X, Y can take only the two values $\pm(1 - x^2)^{1/2}$ (with equal probability), so Y is not independent of the value of X.

The mean is always additive, and the variance is additive for independent (or uncorrelated) RVs:

$$E[X + Y] = E[X] + E[Y] \qquad (\overline{X + Y} = \overline{X} + \overline{Y}) \tag{2.3.3}$$

$$\begin{aligned}
\mathrm{var}[X + Y] &= E[(X + Y - \overline{X + Y})^2] \\
&= E[(X - \overline{X})^2] + 2E[(X - \overline{X})(Y - \overline{Y})] + E[(Y - \overline{Y})^2] \\
&= \mathrm{var}[X] + \mathrm{var}[Y] + 2\mathrm{cov}[X, Y] \qquad (2.3.4) \\
&= \mathrm{var}[X] + \mathrm{var}[Y] \qquad \text{if } \mathrm{cov}[X, Y] = 0. \qquad (2.3.5)
\end{aligned}$$

2.4 Sample mean and standard deviation

Given N independent samples $x_1, x_2, ..., x_N$ of a random variable X, we can estimate basic statistics of X:

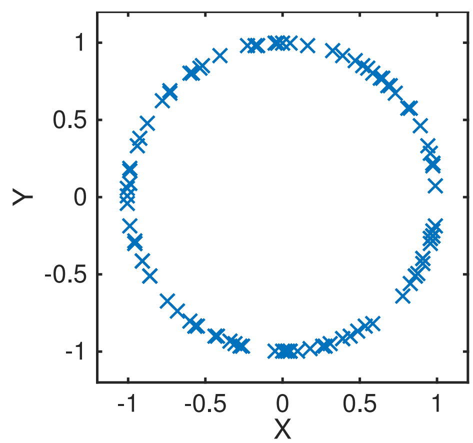

Figure 2: 100 samples of two RVs X and Y which are uncorrelated but dependent

Sample mean

$$\overline{x} = \frac{1}{N}\sum_{j=1}^{N} x_j \tag{2.4.1}$$

The sample mean is an *estimator* of the true mean $\overline{X}$ of X. We will quantify the accuracy of this estimator vs. N later. For now, we note that the sample mean is an *unbiased* estimator of $\overline{X}$, i. e., $E[\overline{x}] = \overline{X}$.

Sample standard deviation $\sigma(x)$ We calculate the variance of the x_j about the *sample* mean $\overline{x}$. Computing the mean from the sample reduces the *effective sample size* (often called the *degrees of freedom* or DOF) by one to $N - 1$:

$$\sigma^2(x) = \mathrm{var}(x) = \frac{1}{N-1}\sum_{j=1}^{N}(x_j - \overline{x})^2 \tag{2.4.2}$$

If the samples are not independent, the effective sample size must be adjusted (Lecture 4). Otherwise $\sigma^2(x)$ is an unbiased estimator of the true variance σ_X^2 of X.

Sample covariance and correlation coefficient between independent samples x_j of RVX and corresponding samples y_j of RV Y:

$$\mathrm{cov}(x, y) = \frac{1}{N-1}\sum_{j=1}^{N}(x_j - \overline{x})(y_j - \overline{y}) \tag{2.4.3}$$

(which is an unbiased estimator of the true covariance between X and Y);

$$R(x, y) = \frac{\mathrm{cov}(x, y)}{\sigma(x)\sigma(y)} \tag{2.4.4}$$

(*not* usually an unbiased estimator of the true correlation coefficient R_{XY}.)

2.4.1 Matlab for sample statistics

If we arrange the x_j into a column vector $\mathbf{x}$:

mean(x) Sample mean.

std(x), var(x) Unbiased standard deviation and variance estimators.

For an array $\mathbf{X}$ these are calculated along the first dimension (the column dimension of a matrix) unless specified otherwise with an optional argument. To get the mean of an array use **mean(X(:))**, i. e. reshape the array into a single vector.

cov(x,y), corrcoef(x,y) Given two column data vectors $\mathbf{x}$ and $\mathbf{y}$, these return 2x2 matrices whose off-diagonal (2,1) and (1,2) elements are the sample covariance (or correlation coefficient).

cov(X), corrcoef(X) Let $\mathbf{X}$ be a $K \times N$ data array whose K columns $\mathbf{x}_k$ correspond to different variables, so that X_{nk} is the n'th sample of variable k. Then these functions return $K \times K$ matrices whose (k, l) entry is the sample covariance $\mathrm{cov}[\mathbf{x}_k, \mathbf{x}_l]$ (or correlation coefficient).