# Review of Probability Theory II

January 29-31, 2008

## 1 Expectation

If the sample space $\Omega = \{\omega_1, \omega_2, \ldots\}$ is countable and $g$ is a real-valued function, then we define the **expected value** or the **expectation** of a function $f$ of $X$ by

$$Eg(X) = \sum_i g(X(\omega_i))P\{\omega_i\}.$$

To create a formula for discrete random variables, write $R(x)$ for the set of $\omega$ so that $X(\omega) = x$, then the sum above can be written

$$\begin{aligned} Eg(X) &= \sum_x \sum_{\omega \in R(x)} g(X(\omega))P\{\omega\} = \sum_x \sum_{\omega \in R(x)} g(x)P\{\omega\} \\ &= \sum_x g(x) \sum_{\omega \in R(x)} P\{\omega\} = \sum_x g(x)P\{\omega; X(\omega) = x\} \\ &= \sum_x g(x)P\{X = x\} = \sum_x g(x)p(x). \end{aligned}$$

provided that the sum converges absolutely. Here $p$ is the mass function for $X$.

For a continuous random variable, with distribution function $F$ and density $f$, choose a small positive value $\Delta x$, and let $\tilde{X}$ be the random variable obtained by rounding the value of $X$ down to the nearest integer multiple of $\Delta x$, then

$$\begin{aligned} Eg(\tilde{X}) = \sum_{\tilde{x}} g(\tilde{x})P\{\tilde{X} = \tilde{x}\} &= \sum_{\tilde{x}} g(\tilde{x})P\{\tilde{x} \le X < \tilde{x} + \Delta x\} = \sum_{\tilde{x}} g(\tilde{x})(F(\tilde{x} + \Delta x) - F(\tilde{x})) \\ &\approx \sum_{\tilde{x}} g(\tilde{x})f(\tilde{x})\Delta x \approx \int_{-\infty}^{+\infty} g(x)f(x)\, dx. \end{aligned}$$

Provided that the integral converges absolutely, these approximations become an equality in the limit as $\Delta x \to 0$.

**Exercise 1.** *Let $X_1$ and $X_2$ be random variables on a countable sample space $\Omega$ having a common state space Let $g_1$ and $g_2$ be two real valued functions on the state space and two numbers $c_1$ and $c_2$. Then*

$$E[c_1 g_1(X_1) + c_2 g_2(X_2)] = c_1 Eg_1(X_1) + c_2 Eg_2(X_2).$$

Several choice for $g$ have special names.

1. If $g(x) = x$, then $\mu = EX$ is call variously the **mean**, and the **first moment.**
2. If $g(x) = x^k$, then $EX^k$ is called the $k$**-th moment**.
3. If $g(x) = (x)_k$, where $(x)_k = x(x-1)\cdots(x-k+1)$, then $E(X)_k$ is called the $k$**-th factorial moment.**
4. If $g(x) = (x-\mu)^k$, then $E(X-\mu)^k$ is called the $k$**-th central moment.**
5. The second central moment $\sigma_X^2 = E(X-\mu)^2$ is called the **variance.** Note that
$$\mathrm{Var}(X) = E(X-\mu)^2 = EX^2 - 2\mu EX + \mu^2 = EX^2 - 2\mu^2 + \mu^2 = EX^2 - \mu^2.$$
6. If $X$ is $\mathbb{R}^d$-valued and $g(x) = e^{i\langle\theta,x\rangle}$, where $\langle\cdot,\cdot\rangle$ is the standard inner product, then $\phi(\theta) = Ee^{i\langle\theta,X\rangle}$ is called the **Fourier transform** or the **characteristic function**.
7. Similarly, if $X$ is $\mathbb{R}^d$-valued and $g(x) = e^{\langle\theta,x\rangle}$, then $m(\theta) = Ee^{\langle\theta,X\rangle}$ is called the **Laplace transform** or the **moment generating function**.
8. If $X$ is $\mathbb{Z}^+$-valued and $g(x) = z^x$, then $\rho(z) = Ez^X = \sum_{x=0}^{\infty} P\{X = x\}z^x$ is called the **(probability) generating function**.

Table of Discrete Random Variables

| random variable | parameters | mean | variance | generating function |
|---|---|---|---|---|
| Bernoulli | $p$ | $p$ | $p(1-p)$ | $(1-p)+pz$ |
| binomial | $n,p$ | $np$ | $np(1-p)$ | $((1-p)+pz)^n$ |
| hypergeometric | $N,n,k$ | $\frac{nk}{N}$ | $\frac{nk}{N}\left(\frac{N-k}{N}\right)\left(\frac{N-n}{N-1}\right)$ | |
| geometric | $p$ | $\frac{1-p}{p}$ | $\frac{1-p}{p^2}$ | $\frac{p}{1-(1-p)z}$ |
| negative binomial | $a,p$ | $a\frac{1-p}{p}$ | $a\frac{1-p}{p^2}$ | $\left(\frac{p}{1-(1-p)z}\right)^a$ |
| Poisson | $\lambda$ | $\lambda$ | $\lambda$ | $\exp(-\lambda(1-z))$ |
| uniform | $a,b$ | $\frac{b-a+1}{2}$ | $\frac{(b-a+1)^2-1}{12}$ | $\frac{z^a}{b-a+1}\frac{1-z^{b-a+1}}{1-z}$ |

Table of Continuous Random Variables

| random variable | parameters | mean | variance | characteristic function |
|---|---|---|---|---|
| beta | $\alpha, \beta$ | $\frac{\alpha}{\alpha+\beta}$ | $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ | $F_{1,1}(a, b; \frac{i\theta}{2\pi})$ |
| Cauchy | $\mu, \sigma^2$ | none | none | $\exp(i\mu\theta - \sigma^2)$ |
| chi-squared | $a$ | $a$ | $2a$ | $\frac{1}{(1-2i\theta)^{a/2}}$ |
| exponential | $\lambda$ | $\frac{1}{\lambda}$ | $\frac{1}{\lambda^2}$ | $\frac{i\lambda}{\theta+i\lambda}$ |
| F | $q, a$ | $\frac{a}{a-2}, a > 2$ | $2a^2\frac{q+a-2}{q(a-4)(a-2)^2}$ | |
| gamma | $\alpha, \beta$ | $\frac{\alpha}{\beta}$ | $\frac{\alpha}{\beta^2}$ | $\left(\frac{i\beta}{\theta+i\beta}\right)^\alpha$ |
| Laplace | $\mu, \sigma$ | $\mu$ | $2\sigma^2$ | $\frac{\exp(i\mu\theta)}{1+\sigma^2\theta^2}$ |
| normal | $\mu, \sigma^2$ | $\mu$ | $\sigma^2$ | $\exp(i\mu\theta - \frac{1}{2}\sigma^2\theta^2)$ |
| Pareto | $\alpha, c$ | $\frac{c\alpha}{\alpha-1}, \alpha > 1$ | $\frac{c^2\alpha}{(\alpha-2)(\alpha-1)^2}$ | |
| t | $a, \mu, \sigma^2$ | $\mu, a > 1$ | $\sigma^2\frac{a}{a-2}, a > 1$ | |
| uniform | $a, b$ | $\frac{a+b}{2}$ | $\frac{(b-a)^2}{12}$ | $-i\frac{\exp(i\theta b)-\exp(i\theta a)}{\theta(b-a)}$ |

## 2 Joint Distributions and Conditioning

A pair of random variables $X_1$ and $X_2$ is called independent if for every pair of events $A_1, A_2$,

$$P\{X_1 \in A_1, X_2 \in A_2\} = P\{X_1 \in A_1\}P\{X_2 \in A_2\}. \qquad (8).$$

For their distribution functions, $F_{X_1}$ and $F_{X_2}$, (8) is equivalent to factoring of the joint distribution function

$$F(x_1, x_2) = F_{X_1}(x_1)F_{X_2}(x_2),$$

to the factoring of joint density for continuous random variables

$$f(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2),$$

to the factoring of the joint mass function for discrete random variables

$$p(x_1, x_2) = p_{X_1}(x_1)p_{X_2}(x_2),$$

and, finally, to the factoring of expectations

$$Eg_1(X_1)g_2(X_2) = Eg_1(X_1)Eg_2(X_2).$$

**Definition 2.** *For a pair of random variables $X_1$ and $X_2$, the* **covariance** *with means $\mu_1$ and $\mu_2$ is defined by*

$$Cov(X_1, X_2) = E(X_1 - \mu_1)(X_2 - \mu_2) = EX_1X_2 - \mu_1\mu_2.$$

*In particular, if $X_1$ and $X_2$ are independent, then $Cov(X_1, X_2) = 0$.*

*The* correlation

$$\rho(X_1, X_2) = \frac{Cov(X_1, X_2)}{\sqrt{Var(X_1)\,Var(X_2)}}.$$

**Exercise 3.** $Var(X_1 + \cdots + X_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} Cov(X_i, X_j)$.

For a pair of jointly continuous random variables, the **marginal density** of $X$ is

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)\ dy.$$

The **conditional density** of $Y$ given $X$ is

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}.$$

The **conditional expectation** is the expectation using the conditional density.

$$E[g(Y)|X = x] = \int_{-\infty}^{+\infty} g(y) f_{Y|X}(y|x)\ dy.$$

Similar expression marginal mass function and conditional mass function, replacing integrals by sums, exists for discrete random variables. The **conditional mass function** of $Y$ given $X$ is

$$p_{Y|X}(y|x) = \frac{p(x,y)}{p_X(x)}.$$

The **conditional expectation** is the expectation using the conditional density.

$$E[g(Y)|X = x] = \sum_y g(y) p_{Y|X}(y|x).$$

# 3 Law of Large Numbers

The law of large numbers states that the long term empirical average of independent random variables $X_1, X_2, \ldots$ having a common distribution function $F$ possessing a mean $\mu$.

In words, we have with probability 1,

$$\bar{X}_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n) = \frac{1}{n} S_n \to \mu \text{ as } n \to \infty.$$

We can defiine the **emprical distribution function**

$$\begin{aligned} \bar{F}_n(x) &= \frac{1}{n}\#(\text{ observations from } X_1, X_2, \ldots, X_n \text{that are less than or equal to } x) \\ &= \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,x]}(X_i). \end{aligned}$$

Then, by the strong law,we have with probability 1,

$$\bar{F}_n(x) \to F(x) \text{ as } n \to \infty.$$

The *Glivenko-Cantelli theorem* states that this convergence is uniform in $x$.

# 4 Central Limit Theorem

For the situation above, we have that

$$\bar{X}_n - \mu \to 0 \text{ as } n \to \infty$$

with probability 1.

The **central limit theorem** states that if we magnify the difference by a factor of $\sqrt{n}$, then we see convergence of the *distributions* to a normal random variable.

**Definition 4.** *A sequence of distribution functions $\{F_n; n \geq 1\}$ is said to* **converge in distribution** *to the distribution function $F$ if*

$$\lim_{n\to\infty} F_n(x) = F(x)$$

*whenever $x$ is a continuity point for $F$.*

**Theorem 5** (Central Limit Theorem)**.** *If the sequence $\{X_n; n \geq 1\}$ introduced above has common variance $\sigma^2$, then*

$$\lim_{n\to\infty} P\left\{\frac{\sqrt{n}}{\sigma}\left(\bar{X}_n - \mu\right) \leq z\right\} = \Phi(z)$$

*where $\Phi$ is the distribution function of a standard normal random variable.*

We often write

$$\frac{\sqrt{n}}{\sigma}\left(\bar{X}_n - \mu\right) = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$