Lecture Notes 1
Brief Review of Basic Probability

1 Probability Review

I assume you know basic probability. Chapters 1-3 are a review. I will assume you have read and understood Chapters 1-3. Here is a very quick review.

1.1 Random Variables

A random variable is a map \( X \) from a set \( \Omega \) (equipped with a probability \( P \)) to \( \mathbb{R} \). We write

\[
P(X \in A) = P(\{\omega \in \Omega : X(\omega) \in A\})
\]

and we write \( X \sim P \) to mean that \( X \) has distribution \( P \). The cumulative distribution function (cdf) of \( X \) is

\[
F_X(x) = F(x) = P(X \leq x).
\]

If \( X \) is discrete, its probability mass function (pmf) is

\[
p_X(x) = p(x) = P(X = x).
\]

If \( X \) is continuous, then its probability density function function (pdf) satisfies

\[
P(X \in A) = \int_A p_X(x)dx = \int_A p(x)dx
\]

and \( p_X(x) = p(x) = F'(x) \). The following are all equivalent:

\[
X \sim P, \quad X \sim F, \quad X \sim p.
\]

Suppose that \( X \sim P \) and \( Y \sim Q \). We say that \( X \) and \( Y \) have the same distribution if \( P(X \in A) = Q(Y \in A) \) for all \( A \). In that case we say that \( X \) and \( Y \) are equal in distribution and we write \( X \overset{d}{=} Y \).

Lemma 1 \( X \overset{d}{=} Y \) if and only if \( F_X(t) = F_Y(t) \) for all \( t \).
1.2 Expected Values

The *mean* or expected value of *g(X)* is

\[
E(g(X)) = \int g(x) dF(x) = \int g(x) dP(x) = \begin{cases} 
\int_{-\infty}^{\infty} g(x)p(x)\,dx & \text{if } X \text{ is continuous} \\
\sum_j g(x_j)p(x_j) & \text{if } X \text{ is discrete}
\end{cases}
\]

Recall that:

1. \(E(\sum_{j=1}^{k} c_j g_j(X)) = \sum_{j=1}^{k} c_j E(g_j(X))\).
2. If \(X_1, \ldots, X_n\) are independent then
   \[E\left(\prod_{i=1}^{n} X_i\right) = \prod_{i} E(X_i).\]
3. We often write \(\mu = E(X)\).
4. \(\sigma^2 = \text{Var}(X) = E((X - \mu)^2)\) is the Variance.
5. \(\text{Var}(X) = E(X^2) - \mu^2\).
6. If \(X_1, \ldots, X_n\) are independent then
   \[\text{Var}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i} a_i^2 \text{Var}(X_i).\]
7. The covariance is
   \[\text{Cov}(X,Y) = E((X - \mu_x)(Y - \mu_y)) = E(XY) - \mu_x \mu_y\]
   and the correlation is \(\rho(X,Y) = \text{Cov}(X,Y)/\sigma_x \sigma_y\). Recall that \(-1 \leq \rho(X,Y) \leq 1\).

---

The **conditional expectation** of \(Y\) given \(X\) is the random variable \(E(Y|X)\) whose value, when \(X = x\) is

\[E(Y|X = x) = \int y \, p(y|x)\,dy\]

where \(p(y|x) = p(x,y)/p(x)\).
The Law of Total Expectation or Law of Iterated Expectation:

\[ E(Y) = \mathbb{E}[\mathbb{E}(Y|X)] = \int \mathbb{E}(Y|X = x)p_X(x)dx. \]

The Law of Total Variance is

\[ \text{Var}(Y) = \text{Var}[\mathbb{E}(Y|X)] + \mathbb{E}[\text{Var}(Y|X)]. \]

The moment generating function (mgf) is

\[ M_X(t) = \mathbb{E}(e^{tX}). \]

If \( M_X(t) = M_Y(t) \) for all \( t \) in an interval around 0 then \( X \overset{d}{=} Y \).

Exercise: show that \( M_X^{(n)}(t) \big|_{t=0} = \mathbb{E}(X^n) \).

### 1.3 Transformations

Let \( Y = g(X) \). Then

\[ F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y) = \int_{A(y)} p_X(x)dx \]

where

\[ A_y = \{ x : g(x) \leq y \}. \]

Then \( p_Y(y) = F_Y'(y) \).

If \( g \) is monotonic, then

\[ p_Y(y) = p_X(h(y)) \left| \frac{dh(y)}{dy} \right| \]

where \( h = g^{-1} \).

**Example 2** Let \( p_X(x) = e^{-x} \) for \( x > 0 \). Hence \( F_X(x) = 1 - e^{-x} \). Let \( Y = g(X) = \log X \).

Then

\[ F_Y(y) = P(Y \leq y) = P(\log(X) \leq y) = P(X \leq e^y) = F_X(e^y) = 1 - e^{-e^y} \]

and \( p_Y(y) = e^y e^{-e^y} \) for \( y \in \mathbb{R} \).

**Example 3** Practice problem. Let \( X \) be uniform on \((-1, 2)\) and let \( Y = X^2 \). Find the density of \( Y \).
Let $Z = g(X,Y)$. For example, $Z = X + Y$ or $Z = X/Y$. Then we find the pdf of $Z$ as follows:

1. For each $z$, find the set $A_z = \{(x,y) : g(x,y) \leq z\}$.

2. Find the CDF

$$F_Z(z) = P(Z \leq z) = P(g(X,Y) \leq z) = P(\{(x,y) : g(x,y) \leq z\}) = \int \int_{A_z} p_{X,Y}(x,y) dx dy.$$ 

3. The pdf is $p_Z(z) = F_Z'(z)$.

**Example 4 Practice problem.** Let $(X,Y)$ be uniform on the unit square. Let $Z = X/Y$. Find the density of $Z$.

### 1.4 Independence

$X$ and $Y$ are *independent* if and only if

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$$

for all $A$ and $B$.

**Theorem 5** Let $(X,Y)$ be a bivariate random vector with $p_{X,Y}(x,y)$. $X$ and $Y$ are independent iff $p_{X,Y}(x,y) = p_X(x)p_Y(y)$.

$X_1, \ldots, X_n$ are independent if and only if

$$\mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n) = \prod_{i=1}^n \mathbb{P}(X_i \in A_i).$$

Thus, $p_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \prod_{i=1}^n p_{X_i}(x_i)$.

If $X_1, \ldots, X_n$ are independent and identically distributed we say they are iid (or that they are a random sample) and we write

$$X_1, \ldots, X_n \sim P \quad \text{or} \quad X_1, \ldots, X_n \sim F \quad \text{or} \quad X_1, \ldots, X_n \sim p.$$

### 1.5 Important Distributions

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

$$p(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}.$$

If $X \in \mathbb{R}^d$ then $X \sim N(\mu, \Sigma)$ if

$$p(x) = \frac{1}{(2\pi)^{d/2}|\Sigma|} \exp\left( -\frac{1}{2} (x-\mu)^T \Sigma^{-1} (x-\mu) \right).$$
$X \sim \chi^2_p$ if $X = \sum_{j=1}^p Z_j^2$ where $Z_1, \ldots, Z_p \sim N(0, 1)$.

$X \sim \text{Bernoulli}(\theta)$ if $\mathbb{P}(X = 1) = \theta$ and $\mathbb{P}(X = 0) = 1 - \theta$ and hence

$$p(x) = \theta^x (1 - \theta)^{1-x} \quad x = 0, 1.$$  

$X \sim \text{Binomial}(\theta)$ if

$$p(x) = \mathbb{P}(X = x) = \binom{n}{x} \theta^x (1 - \theta)^{n-x} \quad x \in \{0, \ldots, n\}.$$  

$X \sim \text{Uniform}(0, \theta)$ if

$$p(x) = \mathbb{I}(0 \leq x \leq \theta) / \theta.$$  

1.6 Sample Mean and Variance

The sample mean is

$$\bar{X} = \frac{1}{n} \sum_i X_i,$$

and the sample variance is

$$S^2 = \frac{1}{n-1} \sum_i (X_i - \bar{X})^2.$$  

Let $X_1, \ldots, X_n$ be iid with $\mu = \mathbb{E}(X_i) = \mu$ and $\sigma^2 = \text{Var}(X_i) = \sigma^2$. Then

$$\mathbb{E}(\bar{X}) = \mu, \quad \text{Var}(\bar{X}) = \frac{\sigma^2}{n}, \quad \mathbb{E}(S^2) = \sigma^2.$$  

**Theorem 6** If $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ then

(a) $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

(b) $(n-1) S^2 / \sigma^2 \sim \chi^2_{n-1}$

(c) $\bar{X}$ and $S^2$ are independent

1.7 Delta Method

If $X \sim N(\mu, \sigma^2)$, $Y = g(X)$ and $\sigma^2$ is small then

$$Y \approx N(g(\mu), \sigma^2 (g'(\mu))^2).$$  

To see this, note that

$$Y = g(X) = g(\mu) + (X - \mu) g'(\mu) + \frac{(X - \mu)^2}{2} g''(\xi)$$
for some ξ. Now \(E((X - \mu)^2) = \sigma^2\) which we are assuming is small and so

\[Y = g(X) \approx g(\mu) + (X - \mu)g'(\mu).\]

Thus

\[E(Y) \approx g(\mu), \quad \text{Var}(Y) \approx (g'(\mu))^2\sigma^2.\]

Hence,

\[g(X) \approx N \left(g(\mu), (g'(\mu))^2\sigma^2\right).\]
Appendix: Common Distributions

Discrete

Uniform
- $X \sim U(1, \ldots, N)$
- $X$ takes values $x = 1, 2, \ldots, N$
- $P(X = x) = \frac{1}{N}$
- $\mathbb{E}(X) = \sum_x x P(X = x) = \sum_x x \frac{1}{N} = \frac{1}{N} \frac{N(N+1)}{2} = \frac{(N+1)}{2}$
- $\mathbb{E}(X^2) = \sum_x x^2 P(X = x) = \sum_x x^2 \frac{1}{N} = \frac{1}{N} \frac{N(N+1)(2N+1)}{6}$

Binomial
- $X \sim \text{Bin}(n, p)$
- $X$ takes values $x = 0, 1, \ldots, n$
- $P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$

Hypergeometric
- $X \sim \text{Hypergeometric}(N, M, K)$
- $P(X = x) = \frac{\binom{M}{x} \binom{N-M}{K-x}}{\binom{N}{K}}$

Geometric
- $X \sim \text{Geom}(p)$
- $P(X = x) = (1 - p)^{x-1} p, \ x = 1, 2, \ldots$
- $\mathbb{E}(X) = \sum_x x (1 - p)^{x-1} p = p \sum_x \frac{d}{dp} ((1 - p)^x) = p \frac{p}{p^2} = \frac{1}{p}$

Poisson
- $X \sim \text{Poisson}(\lambda)$
- $P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}, \ x = 0, 1, 2, \ldots$
- $\mathbb{E}(X) = \text{Var}(X) = \lambda$
- $\mathbb{E}(X) = \sum_{x=0}^{\infty} e^{tx} \frac{e^{-\lambda} \lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t-1)}$.  

$7$
\[ E(X) = \lambda e^t e^{\lambda (e^t - 1)} |_{t=0} = \lambda. \]

- Use mgf to show: if \( X_1 \sim \text{Poisson}(\lambda_1), X_2 \sim \text{Poisson}(\lambda_2), \) independent then \( Y = X_1 + X_2 \sim \text{Poisson}(\lambda_1 + \lambda_2). \)

Multinomial Distribution

The multivariate version of a Binomial is called a Multinomial. Consider drawing a ball from an urn with \( k \) different colors labeled “color 1, color 2, \ldots, \) color \( k. \) Let \( p = (p_1, p_2, \ldots, p_k) \) where \( \sum_j p_j = 1 \) and \( p_j \) is the probability of drawing color \( j. \) Draw \( n \) balls from the urn (independently and with replacement) and let \( X = (X_1, X_2, \ldots, X_k) \) be the count of the number of balls of each color drawn. We say that \( X \) has a Multinomial \( (n, p) \) distribution. The pdf is

\[
p(x) = \binom{n}{x_1, \ldots, x_k} p_1^{x_1} \cdots p_k^{x_k}.
\]

Continuous Distributions

Normal

- \( X \sim N(\mu, \sigma^2) \)
- \( p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{1}{2\sigma^2} (x - \mu)^2 \right\}, \) \( x \in \mathcal{R} \)
- mgf \( M_X(t) = \exp\{\mu t + \sigma^2 t^2/2\}. \)
- \( E(X) = \mu \)
- \( \text{Var}(X) = \sigma^2. \)
- e.g., If \( Z \sim N(0, 1) \) and \( X = \mu + \sigma Z, \) then \( X \sim N(\mu, \sigma^2). \) Show this...

Proof.

\[
M_X(t) = E(e^{tx}) = E(e^{t(\mu + \sigma Z)}) = e^{t\mu} E(e^{t\sigma Z})
= e^{t\mu} M_Z(t\sigma) = e^{t\mu} e^{(t\sigma)^2/2} = e^{t\mu + t^2\sigma^2/2}
\]

which is the mgf of a \( N(\mu, \sigma^2). \)
Alternative proof:

\[ F_X(x) = P(X \leq x) = P(\mu + \sigma Z \leq x) = P\left(Z \leq \frac{x - \mu}{\sigma}\right) \]

\[ = F_Z\left(\frac{x - \mu}{\sigma}\right) \]

\[ p_X(x) = F'_X(x) = p_z\left(\frac{x - \mu}{\sigma}\right) \frac{1}{\sigma} \]

\[ = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2} \left(\frac{x - \mu}{\sigma}\right)^2\right\} \frac{1}{\sigma} \]

\[ = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2} \left(\frac{x - \mu}{\sigma}\right)^2\right\} , \]

which is the pdf of a \( N(\mu, \sigma^2) \). \( \square \)

**Gamma**

- \( X \sim \Gamma(\alpha, \beta) \).
- \( p_X(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, \ x \) a positive real.
- \( \Gamma(\alpha) = \int_0^\infty \frac{1}{\beta^\alpha} x^{\alpha-1} e^{-x/\beta} dx. \)
- Important statistical distribution: \( \chi_p^2 = \Gamma(\frac{p}{2}, 2) \).
- \( \chi_p^2 = \sum_{i=1}^p X_i^2 \), where \( X_i \sim N(0, 1) \), iid.

**Exponential**

- \( X \sim \text{exp}(\beta) \)
- \( p_X(x) = \frac{1}{\beta} e^{-x/\beta}, \ x \) a positive real.
- \( \text{exp}(\beta) = \Gamma(1, \beta) \).
- e.g., Used to model waiting time of a Poisson Process. Suppose \( N \) is the number of phone calls in 1 hour and \( N \sim \text{Poisson}(\lambda) \). Let \( T \) be the time between consecutive phone calls, then \( T \sim \text{exp}(1/\lambda) \) and \( E(T) = (1/\lambda) \).
- If \( X_1, \ldots, X_n \) are iid \( \text{exp}(\beta) \), then \( \sum_i X_i \sim \Gamma(n, \beta) \).
- Memoryless Property: If \( X \sim \text{exp}(\beta) \), then

\[ P(X > t + s | X > t) = P(X > s) . \]
Multivariate Normal Distribution

Let \( Y \in \mathbb{R}^d \). Then \( Y \sim N(\mu, \Sigma) \) if

\[
p(y) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left( -\frac{1}{2} (y - \mu)^T \Sigma^{-1} (y - \mu) \right).
\]

Then \( E(Y) = \mu \) and \( \text{cov}(Y) = \Sigma \). The moment generating function is

\[
M(t) = \exp\left( \mu^T t + \frac{t^T \Sigma t}{2} \right).
\]

**Theorem 7**
(a). If \( Y \sim N(\mu, \Sigma) \), then \( E(Y) = \mu \), \( \text{cov}(Y) = \Sigma \).
(b). If \( Y \sim N(\mu, \Sigma) \) and \( c \) is a scalar, then \( cY \sim N(c\mu, c^2 \Sigma) \).
(c). Let \( Y \sim N(\mu, \Sigma) \). If \( A \) is \( p \times n \) and \( b \) is \( p \times 1 \), then \( AY + b \sim N(A\mu + b, A\Sigma A^T) \).

**Theorem 8** Suppose that \( Y \sim N(\mu, \Sigma) \). Let

\[
Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},
\]

where \( Y_1 \) and \( \mu_1 \) are \( p \times 1 \), and \( \Sigma_{11} \) is \( p \times p \).
(a). \( Y_1 \sim N_p(\mu_1, \Sigma_{11}) \), \( Y_2 \sim N_{n-p}(\mu_2, \Sigma_{22}) \).
(b). \( Y_1 \) and \( Y_2 \) are independent if and only if \( \Sigma_{12} = 0 \).
(c). If \( \Sigma_{22} > 0 \), then the condition distribution of \( Y_1 \) given \( Y_2 \) is

\[
Y_1|Y_2 \sim N_p(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(Y_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}).
\]

**Lemma 9** Let \( Y \sim N(\mu, \sigma^2 I) \), where \( Y^T = (Y_1, \ldots, Y_n), \mu^T = (\mu_1, \ldots, \mu_n) \) and \( \sigma^2 > 0 \) is a scalar. Then the \( Y_i \) are independent, \( Y_i \sim N_1(\mu, \sigma^2) \) and

\[
\frac{||Y||^2}{\sigma^2} = \frac{Y^TY}{\sigma^2} \sim \chi^2_n\left( \frac{\mu^T \mu}{\sigma^2} \right).
\]

**Theorem 10** Let \( Y \sim N(\mu, \Sigma) \). Then:
(a). \( Y^T \Sigma^{-1} Y \sim \chi^2_n(\mu^T \Sigma^{-1} \mu) \).
(b). \( (Y - \mu)^T \Sigma^{-1} (Y - \mu) \sim \chi^2_n(0) \).