# ECE313FA24Note

AutuEnd

December 9, 2024

# 1 Foundations

## 1.1 probability space

### triplet

It's a triplet $(\Omega, \mathcal{F}, P)$

$\Omega$ : A nonempty set, each element $\omega$ of $\Omega$ is called an *outcome* and $\Omega$ is called the *sample space*. The number of $\omega$ is called the cardinality of $\Omega$

$\mathcal{F}$ : Read as *Script F*, a set of all subsets of $\Omega$, also call it *events*.

$P$ : *a probability measure on F*. $P(A)$ is the probability of event $A$ ($A \in \mathcal{F}$).

We use $\complement_A$ or $A^C$ to mean the complement of A.

### Event axioms

$\Omega$ is an event ($\Omega \in \mathcal{F}$).
If $A$ is an event then $A^C$ is an event ($A \in \mathcal{F} \implies A^C \in \mathcal{F}$).
If $A$ and $B$ are events then $A \cup B$ is an event ($A, B \in \mathcal{F} \implies A \cup B \in \mathcal{F}$).

### Probability axioms

$\forall A \in \mathcal{F}, P(A) \geq 0$.
if $A, B \in \mathcal{F}$ and $A$ and $B$ are mutually exclusive , then $P(A \cup B) = P(A) + P(B)$.
$P(\Omega) = 1$

## 1.2 Calculate the size of various sets

**Principle of counting**: If there are $m$ ways to select one variable and $n$ ways to select another varaible, and if these two selections can be made independently, then there is a total of $mn$ ways to make the pair of selections.
**n choose k****: $\binom{n}{k}$

# 2 Discrete-type random variables

## 2.1 A random variable is a real-valued function on $\Omega$

*pmf* (probability mass function): $p_X(u) = P\{X = u\}$
$P\{X \in \{u_1, u_2...\}\} = \sum_i p_X(u_i) = 1$

## 2.2 The mean of a random variable

The *mean* (also called *expectation*) of a random variable $X$ with *pmf* $p_X$ is denoted by $E[X]$ and is defined by $E[X] = \sum_i u_i p_X(u_i)$, where $u_1, u_2, ...$ is the list of possible values of $X$.
The general formula for the mean of a function, $g(X)$, of $X$, is $E[g(X)] = \sum_i g(u_i)p_X(u_i)$.
$E[ag(X) + bh(X) + c] = aE[g(X)] + bE[h(X)] + c$.

## 2.3 The variance and standard deviation of a random variable

The *variance* of a random variable $X$ is a measure of how spread out the pmf of $X$ is. Letting $\mu_X = E[X]$, the variance is defined by: $Var(X) = E[(X - \mu_X)^2] = E[(X - E[X])^2] = E[X^2] - 2\mu_X E[X] + \mu_X^2 = E[X^2] - \mu_X^2$
$E[aX + b] = aE[X] + b$ , $Var(aX + b) = a^2 Var(X)$
The *standardized version* of $X$ is the random variable $\dfrac{X - \mu_X}{\sigma_X}$, and $Var\left(\dfrac{X - \mu_X}{\sigma_X}\right) = 1$

## 2.4 Conditional probabilities

The **conditional probability** of B given A is defined by: $P(B|A) = \begin{cases} \frac{P(AB)}{P(A)} \text{ if } P(A) > 0 \text{undefined if } P(A) = 0 \end{cases}$

#### Mutually independent events

Event $A$ is **independent** of event $B$ if $P(AB) = P(A)P(B)$.
Events $A$,$B$ and $C$ are **pairwise independent** if $P(AB) = P(A)P(B), P(AC) = P(A)P(C), P(BC) = P(B)P(C)$
Events $A$,$B$ and $C$ are **independent** if ther are *pairwise independent* and if $P(ABC) = P(A)P(B)P(C)$

#### Discrete-type indepent random variables

Random variables $X$ and $Y$ are **independent** if any event of the form $X \in A$ is independent of any event of the form $Y \in B$. $(P\{X = i, Y = j\} = p_X(i)p_Y(j))$

## 2.5 Binomial distribution

A random variable $X$ is said to have the **Bernoulli distribution** with parameter $p$, where $0 \leq p \leq 1$, if $p_X(1) = p$ and $p_X(0) = 1 - p$. $E[X] = p$, $Var(X) = E[X^2] - E[X]^2 = p(1 - p)$

Suppose $n$ independent *Bernoulli trials* are conducted, each resulting in a one with probability $p$ and a zero with probability $1 - p$. Let $X$ denote the total number of ones occurring in the $n$ trials. The pmf of $X$ is

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k} \text{ for } 0 \leq k \leq n$$

$E[X] = np$ and $Var[X] = np(1 - p)$

## 2.6 Geometric distribution

Do Bernoulli trials untill the outcome of a trial is one. L denote the number of trials conducted. The pmf of L is:

$$p_L(k) = (1-p)^{k-1} p \text{ for } k \geq 1$$

and $P\{L > k\} = (1 - p)^k$ for $k \geq 0$.
$E[L] = \frac{1}{p}$, $Var[L] = \frac{1-p}{p^2}$

## 2.7 Negative binomial distribution

Let $S_r$ denotes the number of trials required for $r$ ones, and the last trail must be one. Let $n \geq r$, and let $k = n - r$. The event $\{S_r = n\}$ is determined by the outcomes of the ffirst n trials. The event is true iff there are $r - 1$ ones and $k$ zeros in the first $k + r - 1$ trials, and trail $n$ is one. Therefore, the pmf of $S_r$ is given by

$$p(n) = \binom{n-1}{r-1} p^r (1-p)^{n-r} \text{ for } n \geq r$$

$E[S_r] = \frac{r}{p}$, $Var(S_r) = rVar(L_1) = \frac{r(1-p)}{p^2}$

## 2.8 Poisson disttribution

The **Poission probability distribution** with parameter $\lambda > 0$ is the one with pmf $p(k) = \frac{e^{-\lambda}\lambda^k}{k!}$ for $k \geq 0$. It's a good approximation for a binomial distribution with parameters $n$ and $p$, when $n$ is very large, $p$ is very small, and $\lambda = np$.
$E[Y] = Var[Y] = \lambda$
Examples:

- *Radio active emissions in a fixed time interval*: n is the number of uranium atoms in a rock sample, and p is the probability that any particular one of those atoms emits a particle in a one minute period.
- *Incoming phone calls in a fixed time interval*: n is the number of people with cell phones within the access region of one base station, and p is the probability that a given such person will make a call within the next minute.
- *Misspelled words in a document*: n is the number of words in a document and p is the probability that a given word is misspelled.

### 2.9 Maximum likelihood parameter estimation

For a random variable $X$, and that the pmf of $X$ is $p_\theta$, where $\theta$ is a parameter. The probability of k being the observed value for X. The *likelihood* $X = k$ is $p_\theta(k)$. The *maximum likelihood estimate* of $\theta$ for observation $k$, denoted by $\hat{\theta}_{ML}(k)$, is the value of $\theta$ that maximizes the likelihood, $p_\theta(k)$, with respect to $\theta$. (Give k, find $\theta$ (or $p$ in some Bernoulli trials) to make $p_\theta(k)$ biggest, $\hat{\theta}_{ML}(k)$ =the value of $\theta$).

### 2.10 Markov and Chebychev inequalities and confidence intervals

**Markov's inequality**: $P\{Y \geq c\} \leq \dfrac{E[Y]}{c}$

**Chebychev inequalit**: $P\{|X-\mu| \geq d\} \leq \dfrac{\sigma^2}{d^2}$

### 2.11 The law of total probability

$P(E_i|A) = \dfrac{P(AE_i)}{A} = \dfrac{P(A|E_i)P(E_i)}{P(A)}$
$E[X] = \sum_{j=1}^{J} E[X|E_j]P(E_j)$

### 2.12 confidence intervals

$$P\{p \in (\hat{p} - \frac{a}{2\sqrt{n}}, \hat{p} + \frac{a}{2\sqrt{n}})\} \geq 1 - \frac{1}{a^2}$$

### 2.13 Binary hypothesis testing with discrete-type observations

| | $X=0$ | $X=1$ | $X=2$ | $X=3$ |
|---|---|---|---|---|
| $H_1$ | 0.0 | 0.1 | <u>0.3</u> | <u>0.6</u> |
| $H_0$ | <u>0.4</u> | <u>0.3</u> | 0.2 | 0.1 |

**Four possible outcomes**

: Hypothesis $H_0$ is true and $H_0$ is declared.
Hypothesis $H_1$ is true and $H_1$ is declared.
Hypothesis $H_0$ is true and $H_1$ is declared. This is called a *false alarm*.
Hypothesis $H_1$ is true and $H_0$ is declared. This is called a *miss*.
In this example: $p_{false-alarm} = 0.2 + 0.1 = 0.3$, $p_{miss} = 0.0 + 0.1 = 0.1$

### 2.14 Maximum likelihood(ML) decision rule

Define the **likelihood ratio**: $\Lambda(k) = \dfrac{p_1(k)}{p_0(k)}$

An LRT with threshold $\tau$ (For ML, $\tau = 1$) can be written as $\Lambda(X) \begin{cases} > \tau \text{ declare } H_1 \text{ is true} < \tau \text{ declare } H_0 \text{ is true} \end{cases}$
$\Lambda(X) = \tau$? It will metioned in questions(like "Asume that ties are broken in favor of $H_1$").

### 2.15 Maximum a posteriori probability (MAP) decision rule

Define a $(\pi_0, \pi_1)$. $P(H_i, X = k) = \pi_i p_i(k)$. For example, the joint probability matrix ($(\pi_0, \pi_1) = (0.8, 0.2)$) is:

| | $X=0$ | $X=1$ | $X=2$ | $X=3$ |
|---|---|---|---|---|
| $H_1$ | 0.0 | 0.02 | 0.06 | <u>0.12</u> |
| $H_0$ | <u>0.32</u> | <u>0.24</u> | <u>0.16</u> | 0.08 |

The threshold $\tau$ of LRT in MAP is $\dfrac{\pi_0}{\pi_1}$
$p_e = \pi_0 p_{false-alarm} + \pi_1 p_{miss}$ (For ML, It is possible that $p_e \neq 0.5p_{false-alarm} + 0.5p_{miss}$)

### 2.16 Reliability

**Union bound**: $P(A_1 \cup A_2 \cup ... \cup A_m) \leq P(A_1) + P(A_2) + ... + P(A_m)$.

# 3 Continuous-type random vairables

## 3.1 Cumulative Distribution Functions(CDF)

CDF of a discrete-type random variable: $F_X(c) = \sum_{u:u\leq c} p_X(u)$

CDF of a continuous-type random vairable: $F_X(c) = \int_{-\infty}^{c} f_X(u)du$

## 3.2 Continuous-type random variables

$\mu_X = E[X] = \int_{-\infty}^{+\infty} u f_X(u)du$
$Var(X) = E[X^2] - \mu_X^2$

### Law of the Unconscious Statistician

$E[g(X)] = \int_{-\infty}^{\infty} g(u)f_X(u)du$
$Var(g(X)) = \int_{-\infty}^{\infty} g^2(u)f_X(u)du - \left(\int_{-\infty}^{\infty} g(u)f_X(u)du\right)^2$

## 3.3 Uniform distribution

Let $a \leq b$. A random variable $X$ is *uniformly distributed* over the interval $[a, b]$ if

$$f_X(u) = \begin{cases} \frac{1}{b-a} & a \leq u \leq b \\ 0 & else. \end{cases}$$

The pdf and CDF of the uniform distribution over an interval $[a, b]$:

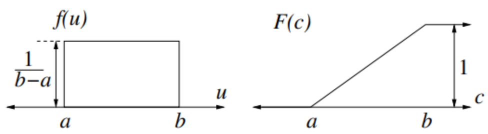


Figure 1: pdf and CDF of the uniform distribution

$E[X] = \dfrac{a+b}{2}, Var[X] = \dfrac{(a-b)^2}{12}$

## 3.4 Exponential distribution

A random variable $T$ has the exponential distribution with parameter $\lambda > 0$ if its pdf is given by

$$f_T(t) = \begin{cases} \lambda e^{-\lambda t} & t \geq 0 \\ 0 & else. \end{cases}$$

and CDF is:

$$F_T(t) = \begin{cases} 1 - e^{-\lambda t} & t \geq 0 \\ 0 & t < 0 \end{cases}, \; F_T^c(t) = \begin{cases} e^{-\lambda t} & t \geq 0 \\ 1 & t < 0 \end{cases}$$

$E[T^n] = \dfrac{n!}{\lambda^n}, Var(T) = \dfrac{1}{\lambda^2}$

**Memoryless property of exponential distribution**: $P(T > s+t | T > s) = \dfrac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P\{T > t\}$.

## 3.5 The Erlang distribution

Let $T_r$ denote the time of the $r^{th}$ count of a Poisson process. Thus, $T_r = U_1 + ... + U_r$.

$$P\{T_r > t\} = \sum_{k=0}^{r-1} \frac{e^{-\lambda t}(\lambda t)^k}{k!}$$

$f_{T_r}(t) = \dfrac{e^{-\lambda t}\lambda^r t^{r-1}}{(r-1)!}$
$E[T_r] = \dfrac{r}{\lambda}, Var[T_r] = \dfrac{r}{\lambda^2}$

### 3.6 The Gaussian (normal) distribution

$$N(\mu,\sigma^2) \quad f(u) = \frac{e^{-\frac{(u-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$$

**standard normal distribution** $N(0,1)$

pdf: $f(u) = \frac{e^{-\frac{(u)^2}{2}}}{\sqrt{2\pi}}$
CDF: $\Phi(u)$, $Q(u) = 1 - \Phi(u)$

### 3.7 ML parameter estimation for continuous-type variables

Example: $\lambda_{ML}(t)$, is the value of $\lambda > 0$ that maximizes $\lambda e^{-\lambda t}$ with respect to $\lambda$, for $t$ fixed. Since $\frac{d(\lambda e^{-\lambda t})}{d\lambda} = (1-\lambda t)e^{-\lambda t}$, the likelihood is increasing in $\lambda$ for $0 \le \lambda \le \frac{1}{t}$, and it is decreasing in $\lambda$ for $\lambda \ge \frac{1}{t}$, so the likelihood is maximized at $\frac{1}{t}$. That is, $\hat{\lambda}_{ML} = \frac{1}{t}$

### 3.8 Function of a random variable

Scope the problem: Get ready. find the CDF of $Y$: $F_Y(c) = P\{Y \le c\} = P\{g(X) \le c\}$ Differentiate $F_Y$ to find its derivative, which is $f_Y$

### 3.9 Binary hypothesis testing with continuous-type observations

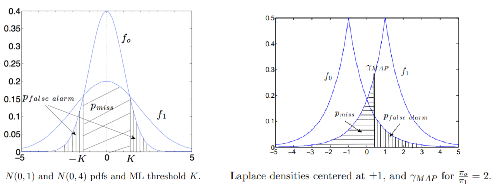


Figure 2: pdf and CDF of the uniform distribution

# 4 Jointly Distributed Random variables

### 4.1 Joint cumulative distribution functions

$F_{X,Y}(u_0, v_0) = P\{X \le u_0, Y \le v_0\}$
For rectangular region $(a,b] \times (c,d]$ in the plane, then $P\{(X,Y) \in \mathbb{R}\} = F_{X,Y}(b,d) - F_{X,Y}(b,c) - F_{X,Y}(a,d) + F_{X,Y}(a,c)$

### 4.2 Joint probability mass functions

$p_X(u) = \sum_j p_{X,Y}(u, v_j), p_Y(v) = \sum_i p_{X,Y}(u_i, v).$

$p_{Y|X}(v|u_o) = P(Y = v|X = u_o) = \frac{p_{X|Y}(u_o, v)}{p_X(u_o)}$

### 4.3 Joint probability density functions

If $g$ is a function on the plane then the expectation of the random vairable $g(X,Y)$ can be computed using LOTUS:

$$E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(u,v) f_{X,Y}(u,v)dudv$$

The pdf of $X$ and $Y$: $\begin{cases} f_X(u) = \int_{-\infty}^{\infty} f_{X,Y}(u,v)dv \\ f_Y(v) = \int_{-\infty}^{\infty} f_{X,Y}(u,v)du \end{cases}$
And the conditional pdf: $f_{Y|X}(v,u_o) = \frac{f_{X,Y}(u_o,v)}{F_X(u_o)}$
The conditional expectation of $Y$ given $X = u$ written as: $E[Y|X=u] = \int_{-\infty}^{\infty} v f_{Y|X}(v,u)dv$

## 4.4 Independence of random variables

Defination 1: Random variables $X$ and $Y$ are defined to be independent if any pair of the form $\{X \in A\}$ and $\{Y \in B\}$, are independent. That is: $P\{X \in A, Y \in B\} = P\{X \in A\}P\{Y \in B\}$, also $F_{X,Y}(u_o, v_o) = F_X(u_o)F_Y(v_o)$

Proposition 2: $X$ and $Y$ are independent if and only if: $\forall u \in \mathbb{R}, \forall v \in \mathbb{R} \rightarrow$ either $f_X(u) = 0$ or $f_{Y|X}(v, u) = f_Y(v)$.

Defination 3: A subset $S$ in $\mathbb{R}^2$ has the swap property if for any two points $(a, b) \in S$ and $(c, d) \in S$, the points $(a, d)$ and $c, b$ are also in $S$.

## 4.5 Joint pdfs of function of random variables

### 4.5.1 Transformation of pdfs under a linear mapping

Suppose $X$ and $Y$ have a joint pdf $f_{X,Y}$, and suppose:

$$\begin{bmatrix} W \\ Z \end{bmatrix} = A \begin{bmatrix} X \\ Y \end{bmatrix}, \text{ where } A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

So that $\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = A \begin{bmatrix} u \\ v \end{bmatrix}$, and $\begin{bmatrix} u \\ v \end{bmatrix} = A^{-1} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$

The $\begin{bmatrix} W \\ Z \end{bmatrix}$ has pdf $f_{W,Z}(\alpha, \beta) = \dfrac{1}{|\det A|} f_{X,Y}\left(A^{-1} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}\right)$

### 4.5.2 Transformation of pdfs under a one-to-one mapping

We assume that $\begin{bmatrix} W \\ Z \end{bmatrix} = g\left(\begin{bmatrix} X \\ Y \end{bmatrix}\right)$, which can be expressed by $\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} g_1(u, v) \\ g_2(u, v) \end{bmatrix}$

The Jacobian of $g$, which we denote be $J$, is the matrix-valued function defined by: $J = J(u, v) = \begin{bmatrix} \frac{\partial g_1}{\partial u} & \frac{\partial g_1}{\partial v} \\ \frac{\partial g_2}{\partial u} & \frac{\partial g_2}{\partial v} \end{bmatrix}$

Then $\begin{bmatrix} W \\ Z \end{bmatrix}$ has the joint pdf given by: $f_{W,Z}(\alpha, \beta) = \dfrac{1}{|\det J|} f_{X,Y}\left(g^{-1}\left(\begin{bmatrix} \alpha \\ \beta \end{bmatrix}\right)\right)$

## 4.6 Correlation and covariance

Let $X$ and $Y$ be random variables with finite second moments. Three important related quantities are:

$$\begin{cases} \text{The correlation: } E[XY] \\ \text{The covariance: } \text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y] \\ \text{The correlation coefficient: } \rho_{X,Y} = \dfrac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \dfrac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \end{cases}$$

Random variables $X$ and $Y$ are called *uncorrelated* if $\text{Cov}(X, Y) = 0$
If $\text{Cov}(X, Y) > 0$, the variables are said to be *positively correlated*.
If $\text{Cov}(X, Y) < 0$, the variables are said to be *negatively correlated*
If $X$ and $Y$ are independent, then $E[XY] = E[X]E[Y]$, which implies that $X$ and $Y$ are uncorrelated.
$\text{Cov}(X + Y, X + Y) = \text{Cov}(X, Y) + \text{Cov}(Y, Y) + \text{Cov}(X, Y) + \text{Cov}(Y, X) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$.
$\text{Cov}(aX + b, cY + d) = ac\text{Cov}(X, Y)$.

## 4.7 Minimum mean square error estimation

### 4.7.1 Constant estimators

The **mean square error(MSE)** for estimating $Y$ by $\delta$ is defined by $E[(Y - \delta)^2]$, if $Y$ is a continuous-type random variable,

$$\text{MSE (for extimation of } Y \text{ by a constant } \delta) = \int_{-\infty}^{+\infty} (y - \delta)^2 f_Y(y) dy$$

.

And we can use the fact that $E[Y - \mu_Y] = 0$ and $\mu_Y - \delta$ is constant to get

$$E[(Y - \delta)^2] = \text{Var}(Y) + (\mu_Y - \delta)^2$$

We can see that the MSE is minimized with respect to $\delta$ if and only if $\delta^* = \mu_Y$, and the minimum possible value is $\text{Var}(Y)$

### 4.7.2 Unconstrained estimators

Conditional expectation indeed gives the optimal estimator:

$$\text{MSE} = E[(Y - g(X))^2] = \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} (v - g(u))^2 f_{Y|X}(v|u) dv \right) f_X(u) du$$

We write $E[Y|X] for g^*(X)$. Which means:

$$g^*(u) = E[Y|X = u] = \int_{-\infty}^{\infty} v f_{Y|X}(v|u) dv$$

And the minimum MSE is:

$$MSE = E[(Y - E[Y|X])^2] = E[Y^2] - E[(E[Y|X])^2]$$

### 4.7.3 Linear estimators

The MSE for the linear estimator $aX + b$ is

$$MSE = E[(Y - (aX + b))^2]$$

$$a^* = \frac{\text{Cov}(Y, X)}{\text{Var}(X)}, b^* = \mu_Y - a\mu_X$$

And the Minimum MSE linear estimator is given by $L^*(X) = \hat{E}[Y|X]$, where

$$\hat{E}[Y|X] = \mu_Y + \frac{\text{Cov}(Y, X)}{\text{Var}(X)}(X - \mu_X) = \mu_Y + \sigma_Y + \rho_{X,Y} \left( \frac{X - \mu_X}{\sigma_X} \right)$$

So:

$$MSE_{\text{min}} = \sigma^2 Y - \frac{\text{Cov}^2(X, Y)}{\text{Var}(X)} = \sigma_Y^2 (1 - \rho_{X,Y}^2) = \sigma_Y^2 - \text{Var}(\hat{E}[Y|X]) = E[Y^2] - E[\hat{E}[Y|X]^2]$$

### 4.7.4 Summary

The following ordering among the three MSEs holds:

$$E[(Y - g^*(X))^2] \qquad \leq \qquad \sigma_Y^2 (1 - \rho_{X,Y}^2) \qquad \leq \qquad \sigma_Y^2$$