Chapter 3, 4
Random Variables

ENCS6161 - Probability and Stochastic Processes

Concordia University

The Notion of a Random Variable

- A random variable X is a *function* that assigns a real number $X(\omega)$ to each outcome ω in the sample space of a random experiment. S is the domain, and $S_X = \{X(\omega) : \omega \in S\}$ is the range of r.v. X
- Example:
  toss a coin $S = \{H, T\}$
  $X(H) = 0, X(T) = 1, S_X = \{0, 1\}$
  $P(H) = P(X = 0) = 0.5, P(T) = P(X = 1) = 0.5$
  measure the temperature, $S = \{\omega | 10 < \omega < 30\}$
  $X(\omega) = \omega$, then $S_X = \{x | 10 < x < 30\}$
  What is $P(X = 25)$?

Cumulative Distribution Function

- $F_X(x) \triangleq P(X \leq x) = P(\omega : X(\omega) \leq x) \quad -\infty < x < \infty$
- Properties of $F_X(x)$
  1. $0 \leq F_X(x) \leq 1$
  2. $\lim_{x \to \infty} F_X(x) = 1$
  3. $\lim_{x \to -\infty} F_X(x) = 0$
  4. $F_X(x)$ is nondecreasing, i.e., if $a < b$ then $F_X(a) \leq F_X(b)$
  5. $F_X(x)$ is continuous from right, i.e., for $h > 0$ $F_X(b) = \lim_{h \to 0} F_X(b + h) = F_X(b^+)$ (see examples below)

Cumulative Distribution Function

- Properties of $F_X(x)$

  6. $P(a < X \leq b) = F_X(b) - F_X(a)$ for $a \leq b$

     Proof: $\{X \leq a\} \cup \{a < X \leq b\} = \{X \leq b\}$

     $\{X \leq a\} \cap \{a < X \leq b\} = ?$

     so $P\{X \leq a\} + P\{a < X \leq b\} = P\{X \leq b\}$

     $\Rightarrow P\{a < X \leq b\} = F_X(b) - F_X(a)$

  7. $P\{X = b\} = F_X(b) - F_X(b^{-})$

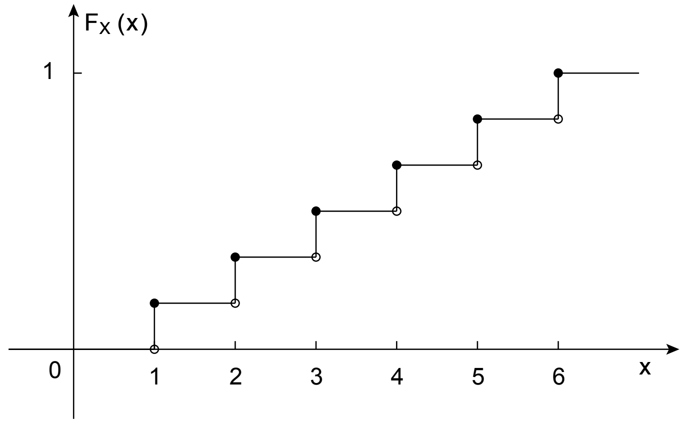
     Proof:

$$
\begin{aligned}
P\{X = b\} &= \lim_{\varepsilon \to 0} P\{b - \varepsilon < X \leq b\} \\
&= F_X(b) - \lim_{\varepsilon \to 0} F_X(b - \epsilon) \\
&= F_X(b) - F_X(b^{-})
\end{aligned}
$$

Cumulative Distribution Function

- Example: roll a dice $S = \{1,2,3,4,5,6\}$, $X(\omega) = \omega$

$$F_X(x) = \begin{cases} 0 & x < 1 \\ \frac{1}{6} & 1 \le x < 2 \\ \frac{2}{6} & 2 \le x < 3 \\ \vdots & \\ \frac{5}{6} & 5 \le x < 6 \\ 1 & x \ge 6 \end{cases}$$

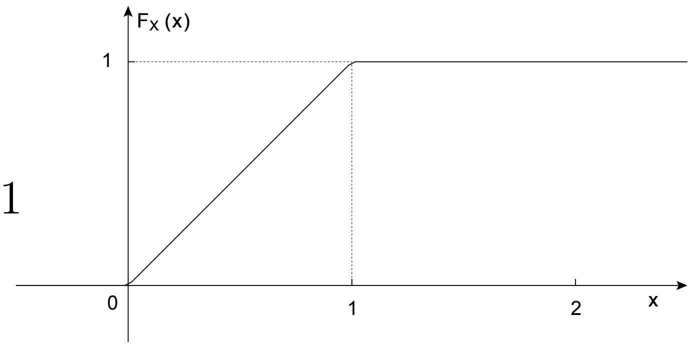


check properties 5,6,7.

Cumulative Distribution Function

- Example: pick a real number between 0 and 1 uniformly

$$F_X(x) = \begin{cases} 0 & x < 0 \\ x/1 & 0 \le x \le 1 \\ 1 & x > 1 \end{cases}$$



- $P\{X = 0.5\} = ?$

Three Types of Random Variables

1. Discrete random variables
   - $S_X = \{x_0, x_1, \cdots\}$ finite or countable
   - Probability mass function (*pmf*)
     $P_X(x_k) = P\{X = x_k\}$
   - $F_X(x) = \sum_k P_X(x_k) u(x - x_k)$ where

$$u(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}$$

   - see the example of rolling a dice
     $F_X(x) = \frac{1}{6}u(x-1) + \frac{1}{6}u(x-2) + \cdots + \frac{1}{6}u(x-6)$

Three Types of Random Variables

2. Continuous random variables $F_x(x)$ is continuous *everywhere*
   $\Rightarrow P\{X = x\} = 0$ for all x
3. Random variables of mixed type
   - $F_X(x) = pF_1(x) + (1-p)F_2(x)$, where $0 < p < 1$
     $F_1(x)$ CDF of a discrete R.V
     $F_2(x)$ CDF of a continuous R.V
   - Example: toss a coin
     if H generate a discrete r.v.
     if T generate a continuous r.v.

Probability density function

- PDF, if it exists is defined as:

  $f_X(x) \triangleq \frac{\mathrm{d}F_X(x)}{\mathrm{d}x}$

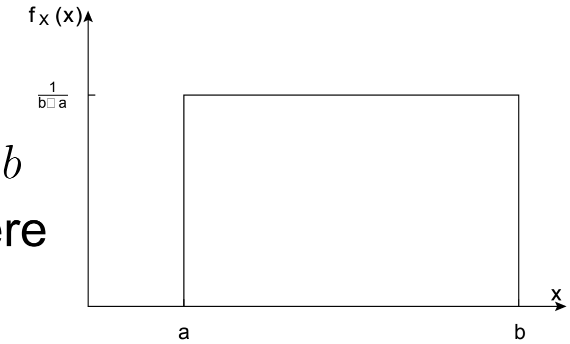
  $f_X(x) \approx \frac{F_X(x+\Delta x) - F_X(x)}{\Delta x} = \frac{P\{x < X \leq x + \Delta x\}}{\Delta x}$ "density"

- Properties of pdf (assume continuous r.v.)
  1. $f_X(x) \geq 0$
  2. $P\{a \leq X \leq b\} = \int_a^b f_X(x)\,\mathrm{d}x$
  3. $F_X(x) = \int_{-\infty}^{x} f_X(t)\,\mathrm{d}t$
  4. $\int_{-\infty}^{+\infty} f_X(t)\,\mathrm{d}t = 1$

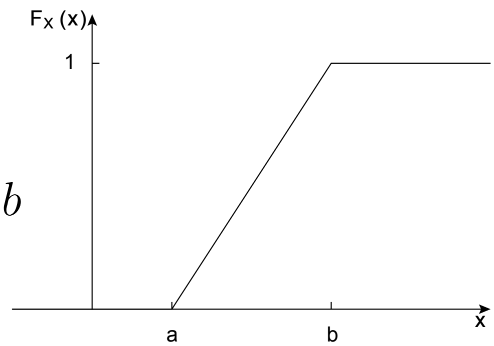
Probability density function

- Example: uniform R.V

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{elsewhere} \end{cases}$$



$$F_X(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$



Probability density function

- Example: if $f_X(x) = ce^{-\alpha|x|}, \; -\infty < x < +\infty$

1. Find constant c

$$\int_{-\infty}^{+\infty} f_X(x)\,\mathrm{d}x = 2\int_{0}^{+\infty} ce^{-\alpha x}\,\mathrm{d}x = \frac{2c}{\alpha} = 1$$

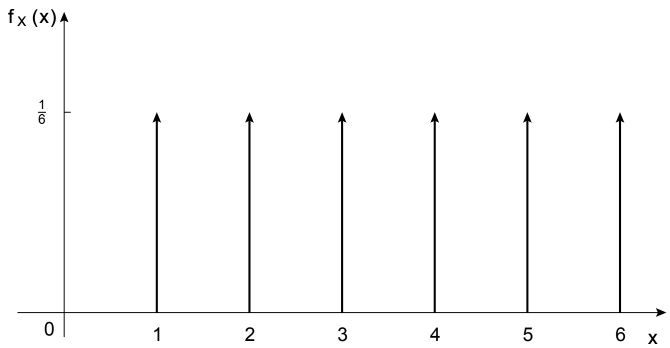
$\Rightarrow c = \frac{\alpha}{2}$

2. Find $P\{|X| < v\}$

$$P\{|X| < v\} = \frac{\alpha}{2}\int_{-v}^{v} e^{-\alpha|x|}\,\mathrm{d}x = 2\,\frac{\alpha}{2}\int_{0}^{v} e^{-\alpha x}\,\mathrm{d}x = 1 - e^{-\alpha v}$$

PDF for a discrete r.v.

- For a discrete r.v. $F_X(x) = \sum_k P_X(x_k)u(x - x_k)$
- We define the *delta function* $\delta(x)$ s.t. $u(x) = \int_{-\infty}^{x} \delta(t)\,dt$ and $\delta(x) = \frac{du(x)}{dx}$
- So pdf for a discrete r.v. $f_X(x) = \sum_k P_X(x_k)\delta(x - x_k)$
- Example:



Conditional CDF and PDF

- Conditional CDF of X given event A

$$F_X(x|A) = P\{X \leq x|A\} = \frac{P\{X \leq x \cap A\}}{P(A)}, \quad P(A) > 0$$

- The conditional pdf

$$f_X(x|A) = \frac{\mathrm{d}F_X(x|A)}{\mathrm{d}x}$$

Conditional CDF and PDF

- Example: the life time X of a machine has CDF $F_X(x)$, find the conditional CDF & PDF given the event $A = \{X > t\}$

$$
\begin{aligned}
F_X(x|X > t) &= P\{X \le x | X > t\} \\
&= \frac{P\{X \le x \cap X > t\}}{P\{X > t\}} \\
&= \begin{cases} 0 & x \le t \\ \frac{F_X(x) - F_X(t)}{1 - F_X(t)} & x > t \end{cases}
\end{aligned}
$$

$$
f_X(x|X > t) = \begin{cases} 0 & x \le t \\ \frac{f_X(x)}{1 - F_X(t)} & X > t \end{cases}
$$

Important Random Variables

- Discrete r.v.
  1. Bernoulli r.v.
  2. Binomial r.v.
  3. Geometric r.v.
  4. Poisson r.v.
- Continuous r.v.
  1. Uniform r.v.
  2. Exponential r.v.
  3. Gaussian (Normal) r.v.

Bernoulli r.v.

- $S_X = \{0, 1\} \quad P_X(0) = 1 - p \quad P_X(1) = p$
  e.g toss a coin
- Let S be a a sample space and $A \subseteq S$ be an event with $P(A) = p$. The *indicator function* of A

  $$I_A(\omega) = \begin{cases} 0 & \text{if } \omega \notin A \\ 1 & \text{if } \omega \in A \end{cases}$$

  I_A is a r.v. since it assigns a number to each outcome of S
- I_A is a Bernoulli r.v.

  $$\begin{aligned} P_{I_A}(0) &= P\{\omega \notin A\} = 1 - P(A) = 1 - p \\ P_{I_A}(1) &= P\{\omega \in A\} = P(A) = p \end{aligned}$$

Binomial r.v.

- Repeat a random experiment n times independently and let X be the number of times that event A with $P(A) = p$ occurs
$$S_X = \{0, 1, \cdots, n\}$$
- Let I_j be the indicator function of the event A in jth trial
$$X = I_1 + I_2 + I_3 \cdots + I_n$$
so X is a sum of Bernoulli r.v.s, where $I_1, I_2, \cdots, I_n$ are *i.i.d* r.v.s (independent identical distribution)
- X is called Binomial r.v.
$$P\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \cdots, n$$

Geometric r.v.

- Count the number X of independent Bernoulli trials until the *first* occurrence of a success. X is called the geometric r.v.

$$S_X = \{1, 2, 3, \cdots\}$$

- Let $p = P(A)$ be the prob of "success" in each Bernoulli trial

$$P\{X = k\} = (1 - p)^{k\Box 1} p, \quad k = 1, 2, \cdots$$

- The geometric r.v. has the memoryless property:

$$P\{X \geq k + j | X > j\} = P\{X \geq k\} \text{ for all } j, k \geq 1$$

prove it by yourself

Poisson r.v.

- Poisson r.v. has the pmf:

$$P\{N = k\} = \frac{\alpha^k}{k!} e^{-\alpha} \quad k = 0, 1, 2, \cdots$$

- A good model for the number of occurrences of an event in a certain time period. α is the avg number of event occurrences in the given time period.
- Example: number of phone calls in 1 hour, number of data packets arrived to a router in 10 mins.
- Poisson prob can be used to approximate a Binomial prob. when n is large and p is small. Let $\alpha = np$

$$P_k = \binom{n}{k} p^k (1 - p)^{n-k} \approx \frac{\alpha^k}{k!} e^{-\alpha} \quad k = 0, 1, \cdots, n$$

Uniform r.v.

- Read on your own.

Exponential r.v.

- $$f_X(x) = \begin{cases} 0 & x < 0 \\ \lambda e^{-\lambda x} & x \geq 0 \end{cases} \qquad F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$
- Good model for the *lifetime* of a device.
- Memoryless: for $t, h > 0$

$$\begin{aligned} P\{X > t + h | X > t\} &= \frac{P\{X > t + h \cap X > t\}}{P\{X > t\}} \\ &= \frac{P\{X > t + h\}}{P\{X > t\}} \\ &= \frac{e^{-\lambda(t+h)}}{e^{-\lambda t}} = e^{-\lambda h} \\ &= P\{X > h\} \end{aligned}$$

Gaussian (Normal) r.v.

- $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$ for $-\infty < x < \infty$
  m, σ are two parameter (will be discussed later)
- CDF:

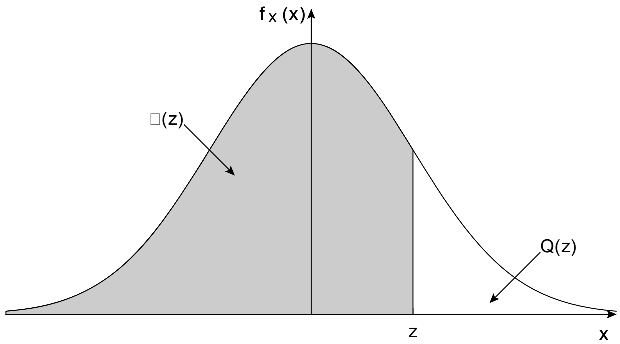
$$\begin{aligned} F_X(x) &= P\{X \le x\} \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-\frac{(x'-m)^2}{2\sigma^2}} \mathrm{d}x' \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-m}{\sigma}} e^{-\frac{t^2}{2}} \mathrm{d}t \quad (t = \frac{x'-m}{\sigma}) \end{aligned}$$

- If we define $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} \mathrm{d}t$ Then

$$F_X = \Phi\left(\frac{x-m}{\sigma}\right)$$

Gaussian (Normal) r.v.

- We usually use $Q(x) = 1 - \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{t^2}{2}} \mathrm{d}t.$



$$\Phi(-x) = Q(x) = 1 - \Phi(x)$$
$$Q(-x) = \Phi(x) = 1 - Q(x)$$

Functions of a random variable

- If X is a r.v., $Y = g(X)$ will also be a r.v.
  The CDF of Y

$$F_Y(y) = P\{Y \leq y\} = P\{g(X) \leq y\}$$

- When $x = g^{-1}(y)$ exists and is unique, the PDF of Y

$$f_Y(y) = \left. \frac{f_X(x)}{|\mathrm{d}y/\mathrm{d}x|} \right|_{x=g^{-1}(y)} = f_X(x) \left| \frac{\mathrm{d}x}{\mathrm{d}y} \right|$$

  Note:

$$f_X(x)\Delta x \approx f_Y(y)\Delta y \Rightarrow f_Y(y) \approx f_X(x) \left| \frac{\Delta x}{\Delta y} \right| \to f_X(x) \left| \frac{\mathrm{d}x}{\mathrm{d}y} \right|$$

- In general, if $y = g(x)$ has n solutions $x_1, x_2, \cdots, x_n$

$$f_Y(y) = \sum_{k=1}^{n} \left. \frac{f_X(x)}{|\mathrm{d}y/\mathrm{d}x|} \right|_{x=x_k} = \sum_{k=1}^{n} f_X(x) \left| \frac{\mathrm{d}x}{\mathrm{d}y} \right|_{x=x_k}$$

Functions of a random variable

- Example: Find the pdf of $Y = aX + b$, in terms of pdf of X. Assume $a \neq 0$ and X is continuous.

$$\begin{aligned} F_Y(y) &= P\{Y \leq y\} \\ &= P\{aX + b \leq y\} \\ &= \begin{cases} P\{X \leq \frac{y-b}{a}\} = F_X(\frac{y-b}{a}) & \text{if } a > 0 \\ P\{X \geq \frac{y-b}{a}\} = 1 - F_X(\frac{y-b}{a}) & \text{if } a < 0 \end{cases} \end{aligned}$$

$$f_Y(y) = \frac{\mathrm{d}F_Y(y)}{\mathrm{d}y} = \begin{cases} \frac{1}{a} f_X(\frac{y-b}{a}) & \text{if } a > 0 \\ -\frac{1}{a} f_X(\frac{y-b}{a}) & \text{if } a < 0 \end{cases} = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

- Example: Let $X \sim N(m, \sigma^2)$ i.e. $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$ if $Y = aX + b$ $(a \neq 0)$ then

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma|a|} e^{-\frac{(y-b-am)^2}{2(a\sigma)^2}}$$

Functions of a random variable

- Example: $Y = X^2$ where X is a continuous r.v.

$$F_Y(y) = P\{Y \le y\} = P\{X^2 \le y\} = \begin{cases} 0 & y < 0 \\ P\{-\sqrt{y} \le X \le \sqrt{y}\} & y \ge 0 \end{cases}$$

$$= \begin{cases} 0 & y < 0 \\ F_X(\sqrt{y}) - F_X(-\sqrt{y}) & y \ge 0 \end{cases}$$

So

$$f_Y(y) = \begin{cases} 0 & y < 0; \\ \frac{f_X(\sqrt{y})}{2\sqrt{y}} + \frac{f_X(-\sqrt{y})}{2\sqrt{y}} & y \ge 0 \end{cases}$$

- If $X \sim N(0; 1)$ then

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \Rightarrow f_Y(y) = \frac{e^{-\frac{y}{2}}}{\sqrt{2\pi y}} \quad \text{for } y \ge 0$$

ch-square r.v. with one degree of freedom

Expected Value

- $E[X] = \int_{-\infty}^{\infty} x f_X(x) \mathrm{d}x$ (may not exist) for continuous r.v.
- $E[X] = \sum_k x_k P_X(x_k)$ for discrete r.v.
- Example: uniform r.v.

$$E[X] = \int_a^b x \frac{1}{b-a} \mathrm{d}x = \frac{b+a}{2}$$

- When $f_X(x)$ is symmetric about $x = m$ then $E[X] = m$. (e.g., Gaussian r.v.)
  $m - x$: odd symmetric about $x = m$
  $(m - x) f_X(x)$: odd symmetric $x = m$
  $\Rightarrow \int_{-\infty}^{\infty} (m - x) f_X(x) \mathrm{d}x = 0$
  $\Rightarrow m = m \int_{-\infty}^{\infty} f_X(x) \mathrm{d}x = \int_{-\infty}^{\infty} x f_X(x) \mathrm{d}x = E[X]$

Expected Value

- Example: the arrival time of packets to a queue has exponential pdf

$$f_X(x) = \lambda e^{-\lambda x} \quad \text{for } x \geq 0$$

$$\begin{aligned} E[X] &= \int_0^\infty x\lambda e^{-\lambda x}\mathrm{d}x = -xe^{-\lambda x}\Big|_0^\infty + \int_0^\infty e^{-\lambda x}\mathrm{d}x \\ &= -\frac{1}{\lambda}e^{-\lambda x}\Big|_0^\infty = \frac{1}{\lambda} \end{aligned}$$

Expected Value

- For a function $Y = g(X)$ of r.v. X

$$E[Y] = \int_{-\infty}^{\infty} g(x) f_X(x) \mathrm{d}x$$

- Example: $Y = a\cos(\omega t + \phi)$ where ϕ is uniformly distributed in $[0, 2\pi]$. Find $E[Y]$ and $E[Y^2]$.

$$E[Y] = \int_0^{2\pi} \frac{1}{2\pi} a\cos(\omega t + \phi)\mathrm{d}\phi = -\frac{a}{2\pi}\sin(\omega t + \phi)\Big|_0^{2\pi} = 0$$

$$E[Y^2] = E[a^2\cos^2(\omega t + \phi)] = E\left[\frac{a^2}{2} + \frac{a^2}{2}\cos(2\omega t + 2\phi)\right] = \frac{a^2}{2}$$

Variance of a random variable

- $\text{Var}[X] = E[(X - E[X])^2] = E[X^2] - E[X]^2$
- Example: uniform r.v.

$$\text{Var}(X) = \int_a^b \frac{1}{b-a}\left(x - \frac{a+b}{2}\right)^2 \mathrm{d}x$$

Let $y = x - \frac{a+b}{2}$

$$\text{Var}(X) = \frac{1}{b-a}\int_{-\frac{b-a}{2}}^{\frac{b-a}{2}} y^2 \mathrm{d}y = \frac{(b-a)^2}{12}$$

Variance of a random variable

- Example: For a Gaussian r.v. $X \sim N(m, \sigma^2)$

$$
\begin{aligned}
\text{Var}(X) &= \int_{-\infty}^{\infty} (x-m)^2 f_X(x)\,\mathrm{d}x \\
&= \int_{-\infty}^{\infty} (x-m)^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}\,\mathrm{d}x \\
&= \int_{-\infty}^{\infty} -(x-m)\frac{\sigma}{\sqrt{2\pi}}\,\mathrm{d}\left(e^{-\frac{(x-m)^2}{2\sigma^2}}\right) \\
&= \left. -(x-m)\frac{\sigma}{\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} \right|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \frac{\sigma}{\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}\,\mathrm{d}x \\
&= \sigma^2 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}\,\mathrm{d}x = \sigma^2
\end{aligned}
$$

Variance of a random variable

- Some properties of variance

$$\begin{aligned} \text{Var}[C] &= 0 \\ \text{Var}[X+C] &= \text{Var}[X] \\ \text{Var}[CX] &= C^2\text{Var}[X] \end{aligned}$$

- nth moment of a random variable

$$E[X^n] \triangleq \int_{-\infty}^{\infty} x^n f_X(x)\,\mathrm{d}x$$

Markov Inequality

- CDF,PDF $\Rightarrow u, \sigma^2$, how to $u, \sigma^2 \overset{?}{\Rightarrow}$ CDF PDF
- The Markov inequality: For non-negative r.v. X

$$P[X \geq a] \leq \frac{E[X]}{a} \quad \text{for } a > 0$$

Proof:

$$\begin{aligned} E[X] &= \int_0^\infty x f_X(x) \mathrm{d}x = \int_0^a x f_X(x) \mathrm{d}x + \int_a^\infty x f_X(x) \mathrm{d}x \\ &\geq \int_a^\infty x f_X(x) \mathrm{d}x \geq \int_a^\infty a f_X(x) \mathrm{d}x = aP[X \geq a] \end{aligned}$$

$$\Rightarrow P[X \geq a] \leq \frac{E[X]}{a}$$

Chebyshev Inequality

- Chebyshev inequality

$$P\{|x-m| \geq a\} \leq \frac{\sigma^2}{a^2}$$

Proof:

$$\begin{aligned} P\{|x-m| \geq a\} &= P\{(x-m)^2 \geq a^2\} \\ &\leq \frac{E[(x-m)^2]}{a^2} = \frac{\sigma^2}{a^2} \\ &\uparrow \text{ Markov inequality} \end{aligned}$$

Transform Methods

- The characteristic function

$$\Phi_X(\omega) = E[e^{j\omega X}] = \int_{-\infty}^{\infty} f_X(x)\, e^{j\omega x} \mathrm{d}x \quad j = \sqrt{-1}$$

  Fourier transform of $f_X(x)$ (with a reversal in the sign of exponent)

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_X(\omega)\, e^{-j\omega x} \mathrm{d}x$$

- Example: Exponential r.v., $f_X(x) = \lambda\, e^{-\lambda x}, x \geq 0$

$$\Phi_X(\omega) = \int_0^{\infty} \lambda\, e^{-\lambda x}\, e^{j\omega x} \mathrm{d}x = \int_0^{\infty} \lambda\, e^{-(\lambda - j\omega)x} \mathrm{d}x = \frac{\lambda}{\lambda - j\omega}$$

Characteristic Function

- If X is a discrete r.v.,

$$\Phi_X(\omega) = \sum_k P_X(x_k)\, e^{j\omega x_k}$$

If x_k are integer-valued,

$$\Phi_X(\omega) = \sum_{k=-\infty}^{\infty} P_X(k)\, e^{j\omega k}$$

$$P_X(k) = \frac{1}{2\pi} \int_0^{2\pi} \Phi_X(\omega)\, e^{-j\omega k} \mathrm{d}\omega$$

Characteristic Function

- Example: For a geometric r.v.,

$$P_X(k) = p(1-p)^k, \quad k = 0, 1, \cdots$$

$$\begin{aligned}\Phi_X(\omega) &= \sum_{k=0}^{\infty} p(1-p)^k e^{j\omega k} \\ &= p \sum_{k=0}^{\infty} [(1-p)e^{j\omega}]^k \\ &= \frac{p}{1-(1-p)e^{j\omega}}\end{aligned}$$

Characteristic Function

- Moment Throrem:

$$E[X^n] = \frac{1}{j^n} \frac{d^n}{d\omega^n} \Phi_X(\omega) \Big|_{\omega=0}$$

Proof:

$$\Phi_X(\omega) = \int_{-\infty}^{\infty} f_X(x)\, e^{j\omega x} dx = \int_{-\infty}^{\infty} f_X(x)\left(1 + j\omega x + \frac{(j\omega x)^2}{2!} + \cdots\right)dx$$

$$= 1 + j\omega E[X] + \frac{(j\omega)^2}{2!} E[X^2] + \cdots$$

so,

$$\Phi_X(\omega)|_{\omega=0} = 1 \qquad (\text{note: } \Phi_X(0) = E[e^0] = E[1] = 1)$$

$$\frac{d}{d\omega}\Phi_X(\omega)|_{\omega=0} = jE[X] \qquad \Longrightarrow E[X] = \frac{1}{j}\frac{d}{d\omega}\Phi_X(\omega)|_{\omega=0}$$

$$\vdots$$

$$\frac{d^n}{d\omega^n}\Phi_X(\omega)|_{\omega=0} = j^n E[X^n] \qquad \Longrightarrow E[X^n] = \frac{1}{j^n}\frac{d^n}{d\omega^n}\Phi_X(\omega)|_{\omega=0}$$

Characteristic Function

- Example: For an exponential r.v.,

$$\Phi_X(\omega) = \frac{\lambda}{\lambda - j\omega}$$

so,

$$\frac{\mathrm{d}}{\mathrm{d}\omega}\Phi_X(\omega) = \frac{\lambda j}{(\lambda - j\omega)^2} \Longrightarrow E[X] = \frac{1}{j}\frac{\mathrm{d}}{\mathrm{d}\omega}\Phi_X(\omega)\,|_{\omega=0} = \frac{1}{\lambda}$$

Find $E[X^2]$ and σ^2 by yourself.

Probability Generating Function

- For non-negative, integer-valued r.v. N,

$$G_N(Z) = E[Z^N] = \sum_{k=0}^{\infty} P_N(k) Z^k$$

which is the Z-transform of the pmf.

$$P_N(k) = \frac{1}{k!} \frac{\mathrm{d}^k}{\mathrm{d}Z^k} G_N(Z)|_{Z=0}$$

- To find the first two moment of N :

$$\frac{d}{dZ} G_N(Z)|_{Z=1} = \sum_{k=0}^{\infty} P_N(k) k Z^{k-1}|_{Z=1} = \sum_k k P_N(k) = E[N]$$

$$\frac{d^2}{dZ^2} G_N(Z)|_{Z=1} = \sum_{k=0}^{\infty} P_N(k) k(k-1) Z^{k-2}|_{Z=1} = \sum_k k(k-1) P_N(k)$$

$$= E[N^2] - E[N]$$

Probability Generating Function

- Example: For a Poisson r.v., $P_N(k) = \frac{\alpha^k}{k!} e^{-\alpha}$

$$G_N(Z) = \sum_{k=0}^{\infty} \frac{\alpha^k}{k!} e^{-\alpha} Z^k = e^{-\alpha} \sum_{k=0}^{\infty} \frac{(\alpha Z)^k}{k!} = e^{-\alpha} e^{\alpha Z} = e^{\alpha(Z-1)}$$

$$G'_N(Z) = \alpha e^{\alpha(Z-1)} \quad G''_N(Z) = \alpha^2 e^{\alpha(Z-1)}$$

$$\Longrightarrow E[N] = G'_N(1) = \alpha \quad E[N^2] - E[N] = G''_N(1) = \alpha^2$$

$$\Rightarrow E[N^2] = \alpha^2 + E[N] = \alpha^2 + \alpha$$

$$Var[N] = E[N^2] - E[N]^2 = \alpha^2 + \alpha - \alpha^2 = \alpha$$

Entropy

- Let X be a r.v. with $S_X = \{1, 2, \cdots, k\}$
  Uncertainty or information of $X = k$

$$I(X = k) = \log \frac{1}{P_X(k)} = -\log P_X(k)$$

- Properties:
  1. $P_X(k) \downarrow$ small, $I(X = k) \uparrow$ large, more information
  2. Additivity, if X, Y independent,

$$\begin{aligned} P(X = k, Y = m) &= P(X = k)P(Y = m) \\ I(X = k, Y = m) &= -\log P(X = k, Y = m) \\ &= -\log P(X = k) - \log P(Y = m) \\ &= I(X = k) + I(Y = m) \end{aligned}$$

Entropy

- Example: toss coins

$$I(X = H) = -\log_2 \frac{1}{2} = 1 \quad \text{(1 bit of info)}$$

$$I(X_1 = H, X_2 = T) = -\log_2 \frac{1}{4} = 2 \text{ (bits)}$$

- If the base of log is 2, we call the unit "bits"
  If the base of log is e, we call the unit "nats"

Entropy

- Entropy:

$$H(X) \triangleq E\left[\log \frac{1}{P_X(k)}\right] = -\sum_k P_X(k) \log P_X(k)$$

- For any r.v. X with $S_X = \{1, 2, \cdots, K\}$

$$H(X) \leq \log K$$

with equality iff $P_k = \frac{1}{K},\ k = 1, 2, \cdots, K$
Read the proof by yourself.

Entropy

- Example: r.v. X with $S_X = \{1, 2, 3, 4\}$ and

	1	2	3	4
P	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$

The entropy:

$$H(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{8}\log\frac{1}{8} - \frac{1}{8}\log\frac{1}{8} = \frac{7}{4}$$

Entropy

- Source Coding Theorem: the minimum average number of bits required to encode a source X is $H(X)$.
- Example:

0	$\longrightarrow$	1
10	$\longrightarrow$	2
110	$\longrightarrow$	3
111	$\longrightarrow$	4

Average number of bits
$= 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{8} + 3 \cdot \frac{1}{8} = \frac{7}{4} = H(X)$:

00	$\longrightarrow$	1
01	$\longrightarrow$	2
10	$\longrightarrow$	3
11	$\longrightarrow$	4

Average number of bits
$= 2 > H(X)$:

Differential Entropy

- For a continuous r.v., since $P\{X = x\} = 0 \Rightarrow$ the entropy is infinite.
- Differential entropy for continuous r.v.

$$H(X) \triangleq -\int_{-\infty}^{\infty} f_X(x) \log f_X(x) \mathrm{d}x = E[-\log f_X(x)]$$

- Example: X is uniform in $[a, b]$

$$H_X = -E\left[\log \frac{1}{b-a}\right] = \log(b-a)$$

Differential Entropy

- Example: $X \sim \mathrm{N}(m, \sigma^2)$

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

$$\begin{aligned} H_X &= -E[\log\ f_X(x)] = -E\left[\log \frac{1}{\sqrt{2\pi}\sigma} - \frac{(x-m)^2}{2\sigma^2}\right] \\ &= \frac{1}{2}\log(2\pi\sigma^2) + \frac{1}{2} = \frac{1}{2}\log(2\pi e\sigma^2) \end{aligned}$$