Exercise 5.

5.1.(a) Markov inequality let X be r.v. which takes only nonnegative values.
Then for any $a > 0$,

$$\mathbb{P}\{X \geqslant a\} \leq \frac{E[X]}{a}.$$

Proof let X be continuous with density f.

$$E[X] = \int_0^\infty x f(x) dx$$
$$= \int_0^a x f(x) + \int_a^\infty x f(x) dx$$
$$\geqslant \int_a^{\cancel{x}} x f(x) dx$$
$$\geqslant \int_a^\infty a f(x) dx$$
$$= a \int_a^\infty f(x) dx$$
$$= a \mathbb{P}\{X \geqslant a\}$$

which implies $P\{X \geqslant a\} \leq \dfrac{E[X]}{a}$

2 (Another Proof ~~2~~) $X = X 1_{(X \geqslant a)} + X 1_{(X < a)}$

$$E[X] = E[X 1_{X \geqslant a}] + E[X 1_{X < a}]$$

clearly $E[X 1_{X \geqslant a}] \geqslant E[a 1_{X \geqslant a}]$

So $E[X] \geqslant E[X 1_{X \geqslant a}] \geqslant E[a 1_{X \geqslant a}]$

$$E[X] \geqslant a E[1_{X \geqslant a}] = a \mathbb{P}(X \geqslant a)$$

which implies $\mathbb{P}(X \geqslant a) \leq \dfrac{E[X]}{a}$ $\square$.

5.1(b) (<u>Chebeshev Inequality</u>)

X : r.v. $\ni$ $\mathbb{E}[X] = \mu$, $Var[X] := \mathbb{V}[X] = \sigma^2$

Then for any value $k > 0$,

$$\mathbb{P}\{|X - \mu| \geq k\} \leq \frac{\sigma^2}{k^2}.$$

<u>Pf</u> $(X-\mu)^2$ is a non-negative r.v. so, we can apply Markov inequality (with $a = k^2$) to get

$$\mathbb{P}\{(X-\mu)^2 \geq k^2\} \leq \frac{E[(X-\mu)^2]}{k^2}.$$

Since $(X-\mu)^2 \geq k^2$ iff $|X-\mu| \geq k$, the above is equivalent to

$$P\{|X - \mu| \geq k\} \leq \frac{E[(X-\mu)^2]}{k^2} = \frac{\sigma^2}{k^2}.$$

$\square$.

5.1(c) Jensen inequality

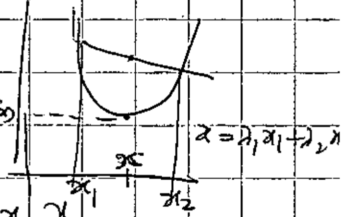
Let X be an (integrable) real-valued r.v. and φ a convex function

Then $$\varphi\left(\mathbb{E}[X]\right) \leq \mathbb{E}[\varphi(X)]$$

(Here the $\mathbb{E}[X]$ denotes the integral of X wrt probability measure $\mathbb{P}$)

We give a proof of this in a simple setting which uses mathematical induction.

Let $\lambda_1, \lambda_2 \geq 0 \ni \lambda_1 + \lambda_2 = 1$.



The convexity of φ implies

$$\varphi(\lambda_1 x_1 + \lambda_2 x_2) \leq \lambda_1 \varphi(x_1) + \lambda_2 \varphi(x_2) \quad \forall x_1, x_2$$

We can easily generalize this:

let $\lambda_1, \lambda_2, \dots \lambda_n \geq 0 \ni \lambda_1 + \lambda_2 + \dots + \lambda_n = 1$.

Then $\varphi(\lambda_1 x_1 + \lambda_2 x_2 + \dots + \lambda_n x_n) \leq \lambda_1 \varphi(x_1) + \dots + \lambda_n \varphi(x_n)$, for any $x_1, \dots, x_n$.

This result can be proved by induction.

By convexity hypothesis, the statement is true for $n=2$.

Suppose it true also for some n, we can prove that it is true for $n+1$.

At least one of λ_i is strictly positive, say λ_1; so by the convexity inequality

$$\varphi\left(\sum_{i=1}^{n+1} \lambda_i x_i\right) = \varphi\left(\lambda_1 x_1 + (1-\lambda_1) \sum_{i=2}^{n+1} \frac{\lambda_i}{1-\lambda_1} x_i\right)$$

$$\leq \lambda_1 \varphi(x_1) + (1-\lambda_1)\, \varphi\left(\sum_{i=2}^{n+1} \frac{\lambda_i}{1-\lambda_1} x_i\right)$$

Since $\sum_{i=2}^{n+1} \frac{\lambda_i}{1-\lambda_1} = 1$, we can apply the induction hypothesis to the last term in the above formula to get

$$\varphi\left(\sum_{i=1}^{n+1} \lambda_i x_i\right) \leq \lambda_1 \varphi(x_1) + \lambda_2 \varphi(x_2) + \dots + \lambda_{n+1} \varphi(x_{n+1})$$

We can write this as

$$\varphi(\mathbb{E}X) \leq \mathbb{E}\,\varphi(X)$$

5.1(c) <u>Jensen inequality - another proof</u>

let φ be convex, and X and $g(X)$ integrable.

Then $\mathbb{E}g(X) \geqslant g(\mathbb{E}X)$.

<u>Pf</u> If φ is convex, then φ lies above all its tangent lines. So for each x_0, there exists $c \ni g(X) \geqslant g(x_0) + c(X - x_0)$.

Letting $x_0 = \mathbb{E}X$ and taking expectations on both sides, we get

$$\underline{\mathbb{E}g(X) \geqslant g(\mathbb{E}X)}$$

$\square$.

5.2. Let $X \sim B(n, p)$. Then X can be approximated by Poisson distribution $P(\lambda)$ when n is large such that $\lambda = np$.

Pf $X \sim B(n, p)$ implies

$$\mathbb{P}\{X = i\} = \frac{n!}{(n-i)!\, i!} p^i (1-p)^{n-i}$$

$$= \frac{n!}{(n-i)!\, i!} \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i}$$

$$= \frac{n(n-1)\cdots(n-i+1)}{n^i} \frac{\lambda^i}{i!} \frac{(1-\lambda/n)^n}{(1-\lambda/n)^i}$$

For large n,

$$\begin{cases} \left(1 - \frac{\lambda}{n}\right)^n \approx e^{-\lambda} \quad \left[\because\ e^x = \lim_{n\to\infty}\left(1 + \frac{x}{n}\right)^n\right] \\ \frac{n(n-1)\cdots(n-i+1)}{n^i} \approx 1 \\ \left(1 - \frac{\lambda}{n}\right)^i \approx 1 \end{cases}$$

Hence, for large n,

$$\mathbb{P}\{X = i\} \approx e^{-\lambda} \cdot \frac{\lambda^i}{i!}$$

i.e. $X \sim P(\lambda)$. $\square$

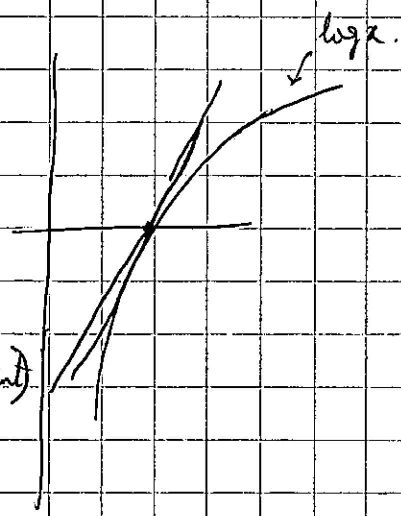
Claim:

For any n, $H(p_1, \dots p_n) \leq \log_2 n$

with equality iff $p_1 = p_2 = \dots = p_n = 1/n$.

Proof

$$\log_2 x = \log_2 e \cdot \log_e x$$



$\log_e x$ is a concave function

(ie $-\log_e x$ is a convex function)

(which geometrically implies that the curve of $\log_e x$ always lies below its tangent)

Drawing the tangent to the curve at $x = 1$,

$$\log_e x \leq x - 1.$$

with equality iff $x = 1$.

So, if $(q_1, \dots q_n)$ is any probability vector, then

$$\log_e (q_k / p_k) \leq (q_k / p_k) - 1$$

with equality iff $q_k = p_k$.

Hence,

$$\sum p_k \log_e (q_k / p_k) \leq \sum q_k - \sum p_k = 0$$

which implies

$$\sum p_k \log q_k \leq \sum p_i \log p_i$$

Putting $q_k = \frac{1}{n}$, $\forall k$,

$$H(p_1, \dots p_n) = -\sum p_i \log_2 p_i \leq \log_2 n$$

with equality as stated (ie $p_1 = p_2 = \dots = p_n = \frac{1}{n}$).

$\square$

5.4. Poisson arrivals implies that interarrival times are exponentially distributed.

Proof:
Consider an arrival process $X \sim P(\lambda)$, ie X is Poisson distributed with mean arrival rate λ.

The probability of exactly n customers arriving during an interval of length t is given by the Poisson law:

$$P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

claim: Poisson arrivals generate an exponential interarrival distribution.
let $a(t)\Delta t$, the probability of an interarrival between t and Δt. it is just the probability of no arrivals for a time t times the probability of one arrival in the infinitesimal interval Δt:

$$a(t)\Delta t = P_0(t)\, P_1(\Delta t)$$

$$P_0(t) = e^{-\lambda t}, \quad P_1(\Delta t) = \lambda \Delta t\, e^{-\lambda \Delta t}$$

In the limit $\Delta t \to 0$ the exponential factor in P_1 tends to unity, so

$$a(t)\,dt = \lambda e^{-\lambda t}\,dt$$

$$\text{So} \quad a(t) = \lambda e^{-\lambda t}.$$

~~ie the arrival proce~~
So, the interarrival times are exponentially distributed:

$$T \sim E(\lambda)$$

5.5 For Poisson arrivals, the number of arrivals in disjoint time intervals are independent.

Poisson process can be defined in three different (but equivalent) ways.

1. Poisson process is a pure birth process (In an infinitesimal interval dt, there may occur only one arrival. This happens with the probability λdt independent of arrivals outside the interval)

2 The # arrivals $N(t)$ in a finite interval of length t obeys the Poisson(λt) distribution, i.e. $N(t) \sim \text{Poisson}(\lambda t)$

$$P\{N(t) = n\} = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

Moreover, the number of arrivals $N(t_1, t_2)$ and $N(t_3, t_4)$ in non-overlapping intervals $(t_1 \le t_2 \le t_3 \le t_4)$ are independent.

3 The interarrival times are independent and obey the $\text{Exp}(\lambda)$ distribution:

$$P(T > t) = e^{-\lambda t}$$

$(2 \rightarrow 3)$

$$\{X > t\} = \{N(t) = 0\} \quad \text{(the events are equivalent)}$$

$$P(X > t) = P(N(t) = 0) = e^{-\lambda t}$$

$$\Rightarrow X \sim \text{Exp}(\lambda)$$

Claim:

5.6 The merging of two independent Poisson arrival streams with parameters λ_1 and λ_2 results in a Poisson stream with parameter $\lambda_1 + \lambda_2$.

Proof: $X_1 \sim P(\lambda_1)$, $X_2 \sim P(\lambda_2)$

merged process $X = X_1 + X_2$

let T be the time until first arrival of the process X.
let T_1 and T_2 be (respectively) the times until the first arrivals of the processes X_1 and X_2.

$$\mathbb{P}(T \geq t) = P(T_1 \geq t)\, P(T_2 \geq t)$$
$$= e^{-\lambda_1 t} \cdot e^{-\lambda_2 t}$$
$$= e^{-t(\lambda_1 + \lambda_2)}.$$

[[In the derivation above, we make use of the fact (a result proved earlier) that Poisson arrivals correspond to exponentially distributed interarrival times.]]

$P(T \geq t) = e^{-t(\lambda_1 + \lambda_2)}$ implies that

$$X \sim P(\lambda_1 + \lambda_2)$$

$\square$.

Corollary It is straightforward to extend the above result to the case of n independent Poisson arrival streams that are merged:

$$X_1 \sim P(\lambda_1), \ \ldots \ X_n \sim P(\lambda_n)$$

$X_1, \ldots X_n$ are independent.

Then $X_1 + X_2 + \ldots + X_n \sim P(\lambda_1 + \ldots + \lambda_n)$

$\square$.

5.7 Consider the Markov chain with state transition matrix

$$P = \begin{pmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{pmatrix}$$

Let π be the steady-state probability vector of P; it satisfies the equation

$$\pi P = \pi \quad \text{ie}$$

$$(\pi_0, \pi_1) \begin{pmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{pmatrix} = (\pi_0 \quad \pi_1)$$

which yields

$$\pi_0 = \alpha \pi_0 + \beta \pi_1$$

$$\pi_1 = (1-\alpha)\pi_0 + (1-\beta)\pi_1$$

and $\pi_0 + \pi_1 = 1$

Solving, $\pi_0 = \dfrac{\beta}{1+\beta-\alpha}$, $\pi_1 = \dfrac{1-\alpha}{1+\beta-\alpha}$

Mathematical Modelling for Computer Networks
Spring 2013

Exercise 5: Due on 19th April 2013.

Write your answers to the questions briefly and clearly. Please bring a printout (or a handwritten copy) of your answers to the class. You may refer to the book Introduction to Probability by Grinstead and Snell (`http://www.dartmouth.edu/~chance/teaching_aids/books_articles/probability_book/book.html`)

1. Markov inequality, Chebechev inequality and Jensen inequality are basic inequalities in probability. State and derive them.
2. Show that binomial distribution $B(n,p)$ can be approximated by Poisson distribution $P(\lambda)$ when n is large and p is small such that $\lambda = np$
3. Information can be regarded as 'dual' of probability; the information associated with probability p is given by $I(p) = -log\, p$ taken with a suitable base for the logarithm such as $e, 10, 2$. When the base is chosen as 2, the amount of information is given in bits. Given a probability vector $\mathbf{p} = (p_1, \ldots, p_n)$, the entropy of associated with $\mathbf{p}$, denoted by $H(\mathbf{p}) = \mathbf{E}(-log\,(\mathbf{p})) \doteq -\sum_{1}^{n} p_i \, log\, p_i$ where $\mathbf{E}$ is the expectation operator wrt the probability vector $\mathbf{p}$. Show that for an arbitrary probability n-vector $\mathbf{p} = (p_1, p_2, ...p_n)$, the entropy $H(\mathbf{p})$ is maximum when all the component probabilities p_i are equal. Interpret the result.
4. Show that Poisson arrivals implies that interarrival times are exponentially distributed.
5. Show that for Poisson arrivals, the number of arrivals in disjoint time intervals are independent.
6. Show that in the merging of two independent Poisson arrival streams with parameters λ_1 and λ_2, the resulting stream is also Poisson with parameter $\lambda_1 + \lambda_2$.
7. The invariant distribution of (an irreducible) Markov chain is a probability vector π such that $\pi = \pi\mathbf{P}$ where $\mathbf{P}$ is one-step transition matrix of the Markov chain. Calculate π for the 2-state Markov chain

$$\mathbf{P} = \begin{bmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{bmatrix}$$

   Evaluate π_0 for the case $\alpha = 0.7$ and $\beta = 0.7$