Problem p.56 #2 $p = 8 \times 10^{-4}$, what is prob. of at least 1 error in 1024 bits if independent?

Ch. 2 Discrete Random Variables

Random variables assign a numerical value to each element of the sample space. The result is a new probability space that is easier to work with mathematically.

<u>Def</u> (partial): A random variable X on a sample space S is a function $X: S \to \mathbb{R}$ that assigns a real number $X(s)$ to each $s \in S$.

Ex. Consider the probability space for a Bernoulli trial. $S = \{s, f\}$. Define X by $X(s) = 1$ $X(f) = 0$.

Ex. Consider a sequence of 4 independent, identically distributed Bernoulli trials. A natural sample space S would be

$\{ c_1 c_2 c_3 c_4 \mid c_i \in \{s, f\} \quad 1 \leq i \leq 4 \}$.

The function $X: S \to \mathbb{N}$ by $X(c_1 c_2 c_3 c_4) =$ # of s's in the sequence, f.ex $X(fssf) = 2$, $X(ffsf) = 1$, etc, is a very common random variable.

A random variable X which takes on a finite or countable set of values, ie $\{x \in \mathbb{R} \mid \exists s \in S\,(X(s)=x)\}$ is countable or finite, is called a discrete random variable.

Given a discrete random variable X on a probability space $(S, \mathcal{F}, P)$, set

$$A_x = \{s \in S \mid X(s) = x\},$$

the inverse image of x under X.

Note that the collection $\{A_x \mid \exists s \in S\,(X(s)=x)\}$ is a partition of S.

(For a function $X : S \to \mathbb{R}$ with discrete image to be a random variable on $(S, \mathcal{F}, P)$, we require that each $A_x \in \mathcal{F}$.)

We often use the shorthand

$$A_x = [X = x].$$

Ex. Consider n iid Bernoulli trials with probability of success p on each trial. The random variable X that assigns to each sequence of S's and F's the number of successes is the basis of the binomial distribution.

$$P(X=k) = \binom{n}{k} p^k q^{n-k}$$

2.3

Given a discrete random variable X on $(S, \mathcal{F}, P)$ and A_x as above,

define the <u>probability mass function</u> or pmf or discrete density function on $\mathbb{R}$ by

$$p_X(x) = P(X=x) = P(A_x) \quad \Big(= \sum_{X(s)=x} P(s)$$

if S is discrete$\Big)$

Ex. If X counts the number of successes in 4 iid Bernoulli trials with probability of success .3,

$$X(0) = \binom{4}{0}(.7)^4$$

$$X(1/2) = 0$$

$$X(2) = \binom{4}{2}(.3)^2(.7)^2$$

$$X(57) = 0$$

If X is a discrete random variable, define a triple $(S', \mathcal{F}', P')$ by

$$S' = \{x \in \mathbb{R} \mid x = X(s) \text{ for some } s \in S\}$$

$F' = \mathcal{P}(S')$, the set of all subsets of F'

$$P'(B) = \sum_{x_i \in B} p_X(x_i) \quad \text{for all } B \in F'.$$

Note this is a valid probability space.

This probability space is the one will will use when tracking the probability of events defined by X.

2.4 Distribution Functions

Def. If X is a discrete random variable, the <u>cumulative distribution function</u>, <u>CDF</u>, or <u>distribution function</u> of X is

$F : \mathbb{R} \to [0,1]$ defined by

$$F(t) = P(X \in (-\infty, t]) = \sum_{x \le t} p_X(x)$$
$$= P(X \le t)$$

Why is

$0 \le F(t) \le 1$ for all $t \in \mathbb{R}$?

F monotone increasing?

$\lim_{t \to -\infty} F(t) = 0$? $\lim_{t \to \infty} F(t) = 1$?

Note if $p_X(x_i)$ is nonzero exactly for $x_1 < x_2 < \dots x_n < \dots$, $i = 1, 2, 3 \dots$ then $F(x_{i+1}) = F(x_i) + p_X(x_{i+1})$ and

$F(x) = F(x_i)$ for all $x_i \le x < x_{i+1}$.

Ex.
For 4 Bernoulli trials w. probability of success .3, set X equal to the number of successes.

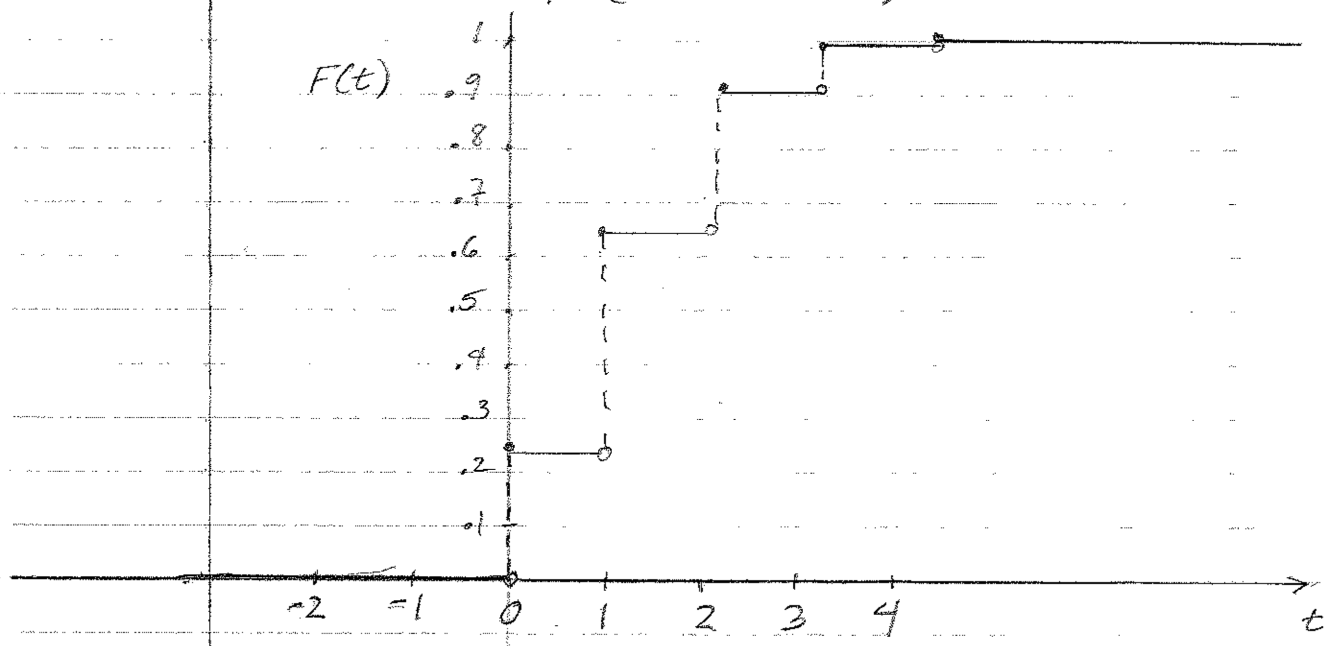
$P(X=0) = .7^4 \approx .24$

$P(X=1) = \binom{4}{1}(.3)(.7^3) \approx .41$, $\quad F(1) \approx .24 + .41 = .65$

$P(X=2) = \binom{4}{2}(.3)^2(.7^2) \approx .26$, $\quad F(2) \approx .65 + .26 = .91$

$P(X=3) = \binom{4}{3}(.3^3)(.7) \approx .08$, $\quad F(3) \approx .91 + .08 = .99$

$P(X=4) = \binom{4}{4}(.3^4) \approx .01$, $\quad F(4) \approx .99 + .01 = 1$



Note that we can construct the pmf from the CDF and vice versa.

2.5 Special Discrete Distributions

These are summarized on p. 777 of the text.

Bernoulli pmf: X is a random variable that takes on the values 0 and 1 only.

pmf: $p_X(0) = q$, $p_X(1) = p$, $p + q = 1$.

CDF: $F(x) = \begin{cases} 0 & x < 0 \\ q & 1 \le x < 1 \\ 1 & x \ge 1 \end{cases}$

Binomial pmf: Y_n models the count of successes in n independent Bernoulli trials, each with probability of success equal to p.

pmf: $p(k) = p(Y_n = k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & 0 \le k \le n \\ 0 & \text{otherwise} \end{cases}$

CDF: $F(t) = \underbrace{B(t; n, p)}_{\text{conventional notation for this function}} = \sum_{i=0}^{\lfloor t \rfloor} \binom{n}{i} p^i (1-p)^{n-i}$

Ex.
Draw n components at random with replacement from a batch of components in which a proportion p are defective. The count of defective components is distributed according to $B(n,p)$.

Ex. Test n methods for improving performance of a system. Report methods that achieve results for which the probability of a result that good or better is $\leq p$ under the original system. Suppose that none of the methods actually alter the performance and that the results of the tests are independent. What is the probability that you will report no improvements?

Simpson's Paradox: You can't be simple-minded when determining relative performance

Undemanding application

	Source A	Source B
good	9	30
failed	1	5
	10% failure	~16.7% failure

Demanding application

	Source A	Source B
good	20	7
failed	20	8
	50%	53%

Which source supplies a better product?

Overall failure rate of source A 21/50.
source B 13/50.

For large n, $C(n,k)$ and so $B(t:n,p)$ can be inconvenient. A couple approximations are available.

Stirling's approximation:

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n \le n! \le \sqrt{2\pi n}\left(\frac{n}{e}\right)^{n+(1/12n)}$$

Normal approximation

$$b(k:n,p) \approx \frac{1}{\sqrt{2\pi np(1-p)}} e^{-(k-np)^2/2npq}$$

Geometric Random Variable

$$p_Z(i) = q^{i-1}p \qquad (p+q=1)$$

$$F_Z(t) = \sum_{i=1}^{\lfloor t \rfloor} p(1-p)^{i-1} = 1-(1-p)^{\lfloor t \rfloor}$$

$$\left(= p\sum_{i=1}^{\lfloor t \rfloor} q^{i-1} = p\,\frac{1-q^{\lfloor t \rfloor}}{1-q} = 1-q^{\lfloor t \rfloor} = 1-(1-p)^{\lfloor t \rfloor}\right)$$

This random variable is motived by counting the number of iid Bernoulli trials required to produce the first success. (The trial resulting in the success is counted)

Record F's on p. 777.

Modified Geometric Random Variable

pmf: $P_Z(i) = q^i p$

$$F_Z(t): \sum_{i=0}^{\lfloor t \rfloor} p(1-p)^i = 1-(1-p)^{\lfloor t \rfloor + 1}$$

This random variable can be generated by counting the number of iid Bernoulli trials up to but not including the first success.

A key property of the geometric random variables is that they are memoryless in the following sense. If Z has a geometric distribution

$$P(Z = n+i \mid Z > n) = P(Z = i) \quad i \in \mathbb{Z}^+$$

"The probability of requiring i additional trials to get a success, given that the first n trials were failures, equals the probability of requiring i trials to get a success from the beginning"

Verification for Geometric RV

$$P(Z=n+i \mid Z>n) = \frac{P(Z=n+i)}{P(Z>n)}$$

$$= \frac{p_Z(n+i)}{1-F_Z(n)} = \frac{pq^{n+i-1}}{1-(1-q^n)} = pq^{i-1} = p_Z(i)$$

(Modified Geometric is similar)

Negative Binomial RV

This family of random variables has 2 parameters, $r \in \mathbb{Z}^+$ and p, $0 \le p \le 1$.

$$p_{T_r}(n) = \binom{n-1}{r-1} p^r (1-p)^{n-r}$$

It models the probability that n repetitions of a sequence of iid Bernoulli trials are required to get r successes.

(Think $n-1$ trials with $r-1$ successes distributed among them, followed by a success)

By the way, though your book contains some statistical tables, greater convenience and accuracy avail. through software. Do you have access to SAS? I use a free open source program called R avail. at http://cran.r-project.org

Poisson RV

The family of Poisson Random Variables has 1 parameter $\alpha > 0$

$$p_\alpha(i) = \frac{e^{-\alpha}\alpha^i}{i!} \qquad i \in \mathbb{N}$$

This important random variable models the number of occurrences of an event in a time interval t if the probability of an occurrence in a short time interval Δt is $\lambda \Delta t$ and the probability of 2 or more occurrences in time Δt is $o(\Delta t)$, aka "small".

Derivation: To estimate the probability of k occurrences in the time interval $(0, t]$, divide $(0, t]$ into n subintervals of equal length and approximate the probability as the probability of k successes in n Bernoulli trials with $p = \lambda\left(\frac{t}{n}\right)$

$$b\left(k; n, \frac{\lambda t}{n}\right) = \binom{n}{k}\left(\frac{\lambda t}{n}\right)^k\left(1 - \frac{\lambda t}{n}\right)^{n-k}$$

The smaller $\frac{t}{n}$, the more accurate the approximation, so consider

$$\lim_{n \to \infty} b\left(k; n, \frac{\lambda t}{n}\right)$$

$$= \lim_{n\to\infty} \cancelto{1}{\frac{(n)}{n}}\cancelto{1}{\frac{(n-1)}{n}}\cdots\cancelto{1}{\frac{(n-k+1)}{n}} \frac{(\lambda t)^k}{k!} \cancelto{1}{\left(1-\frac{\lambda t}{n}\right)^{-k}} \left(1+\frac{\lambda t}{n}\right)^n$$

Invoking differential calculus, note

† $\lim_{n\to\infty} \left(1+\frac{\lambda t}{n}\right)^n = e^{-\lambda t}$,

giving us a probability of $e^{-\lambda t}\frac{(\lambda t)^k}{k!}$ for k events in $[0,t]$. This is the Poisson RV with $\alpha = \lambda t$.

† Calculation of the limit for those interested:

$$\lim_{n\to\infty} \left(1 \mp \frac{\lambda t}{n}\right)^n = \left(\exp\left(\lim_{n\to\infty} \log\left(\left(1-\frac{\lambda t}{n}\right)^{\frac{-n}{\lambda t}}\right)\right)\right)^{-\lambda t}$$

$$= \left(\exp\left(\lim_{n\to\infty}\left(\frac{\log\left(1+\frac{\lambda t}{n}\right)}{-\frac{\lambda t}{n}}\right)\right)\right)^{-\lambda t}$$

$$= \left(\exp\left(\lim_{h\to\infty}\left(\frac{\log(1+h)}{h}\right)\right)\right)^{-\lambda t}$$

$$= \left(\exp\left(\lim_{h\to\infty}\left(\frac{\frac{1}{1+h}}{1}\right)\right)\right)^{-\lambda t}$$ by L'Hôpital's rule

$$= (\exp(1))^{-\lambda t} = e^{-\lambda t}$$

As the derivation indicates, a Poisson RV with $\alpha = np$ can be used to approximate

$$b(k; n, p) \text{ as } \approx e^{-np} \frac{(np)^k}{k!}.$$

One rule of thumb is to avoid this approximation unless $n \geq 20$ and $p \leq .05$

The Poisson RV with $\alpha = \lambda t$ is often used to model arrival rate of jobs in queuing theory for average arrival rate of λ.

Ex. Problem 10, p. 92 reduces to this. If VSLI chips in a system have a failure rate on average of 1 chip every 5 weeks, what is the probability of 3 or more failures in 7 weeks?

The Hypergeometric RV

The hypergeometric random variable has 3 parameters, m, d, n. The pmf is

$$h(k; m, d, n) = \frac{\binom{d}{k}\binom{n-d}{m-k}}{\binom{n}{m}}$$

It models the count of defective items in a sample of size m drawn from a pool of n items of which d are defective.

The Discrete Uniform RV

The random variable X has finite image $\{x_1, x_2, \dots x_n\}$.

$$p_X(x) = \begin{cases} \frac{1}{n} & x \in \{x_1, \dots x_n\} \\ 0 & \text{otherwise} \end{cases}$$

The Constant RV

$$X \equiv c, \quad p_X(x) = \begin{cases} 1 & x = c \\ 0 & \text{otherwise} \end{cases}$$

Indicator RV

This is defined in terms of an event A in a probability space $(S, \mathcal{F}, P)$

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

$$p_{I_A}(1) = P(A)$$

$$p_{I_A}(0) = P(\bar{A})$$

Try #5 p. 91

The probability that a message is transmitted correctly is p. If it is not transitted correctly it is retransmitted. Assuming independence, what is the probability of no resend? Of exactly 2?

p. 92 #9 Assume the number of messages in a time interval t sec. is Poisson distributed with parameter .3t. Compute

a. prob. of exactly 3 messages in 10 sec.

b. prob of at most 20 messages in 20 sec.

c. prob. of betw. 3 & 7 messages in 5 sec.

Skip 2.6

2.7 The Probability Generating Function

Def. If X is a discrete RV with image in $\mathbb{N}$, its probability generating function, PGF is the function:

$$G_X(z) = \sum_{i=0}^{\infty} p_i z^i = p_0 + p_1 z + p_2 z^2 + \ldots$$

where $p_i = P(X=i)$. G_X converges for $|z|<1$, and may converge on a larger interval.

Theorem: If two discrete RV's have the same PGF's then they have the same pmf's.

The usefulness of PGF's comes from this, the fact that information about X can be retrieved from $G_X(0)$, $G_X'(0)$ etc, and from the fact that if X and Y are independent random variables (defined later) then the random variable $X+Y$ has as its PGF the product $G_X(z)G_Y(z)$.

Ex. Poisson:

$$G_X(z) = \sum_{k=0}^{\infty} \frac{\alpha^k}{k!} e^{-\alpha} z^k$$

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

$$= e^{-\alpha} e^{\alpha z} = e^{-\alpha(1-z)}$$

(Don't worry if you didn't recognize the Taylor series for e^x.)

2.8 Discrete Random Vectors

Often we may be interested in the relationships between 2 or more RV's defined on a sample space. For example, if X and Y are RV's giving finishing times for processes executed sequentially,

$X+Y$ is the total time for both processes to finish. If X and Y run concurrently, $\max\{X, Y\}$ is the overall time.

Let $X_1, X_2, \ldots X_r$ be random variables defined on $(S, \mathcal{F}, P)$.

Then the r-tuple $(X_1, X_2, \ldots X_r) = \vec{X}$ is a discrete random vector on $(S, \mathcal{F}, P)$

Def. The joint pmf for a random vector $\vec{X}$ is

$$p_{\vec{X}}(\vec{x}) = P(\vec{X} = \vec{x}) = P(X_1 = x_1 \ \& \ X_2 = x_2 \ \& \ldots X_r = x_r)$$

Note

* $p_{\vec{X}}(\vec{x}) \geq 0$

* $\{\vec{x} \mid p_{\vec{X}}(\vec{x}) \neq 0\}$ is a finite or countably infinite subset of $\mathbb{R}^n$. Denote it by $\{\vec{x}_1, \vec{x}_2, \ldots \vec{x}_k \ldots\}$

* $P(\vec{X} \in A) = \sum_{\vec{x} \in A} p_{\vec{X}}(\vec{x})$

* $\sum_i p_{\vec{X}}(\vec{x}_i) = 1$

In fact, $\vec{X}$ gives rise to a probability space $(S', \mathcal{F}', P')$

$S' = \{\vec{x}_1, \vec{x}_2, \ldots \vec{x}_k \ldots\}$

$\mathcal{F}' = \mathcal{P}(S')$

$P'(A) = \sum_{\vec{x} \in A} p_{\vec{X}}(\vec{x})$.

Then X_i is just projection onto the i^{th} component of S'!

In a similar way, any realvalued function p on $\mathbb{R}^n$ with the '$\star$' properties is the joint pmf of $(S', \mathcal{F}', P')$ as constructed above.

Ex. Suppose X and Y are execution times of 2 modules of a program with the following joint pmf:

	$y=15$	$y=20$	$y=25$	$y=30$ time units
$x=10$	$5/20$	$3/20$	$1/20$	$2/20$
$x=20$	$3/20$	$2/20$	$2/20$	$2/20$

What is the probability, if the modules run sequentially, that the program requires 40 or more time units?

What is the probability that $X=20$?

What is the probability that $Y=25$?

Def. The pmf of one component of a random vector $\vec{X}$ is called its marginal pmf

$$p_{X_i}(x) = \sum_{\substack{\vec{x}_j \in \{\vec{x}_1, \vec{x}_2, \ldots \vec{x}_k \ldots\} \\ \text{and } (\vec{x}_j)_i = x}} p_{\vec{X}}(\vec{x}_j)$$

joint pmf $\to$ marginal pmf often easy

marginal pmf $\to$ joint pmf usually requires additional information

Example: The multinomial distribution can be considered a joint pmf for $X_1, X_2, \ldots X_r$.

Consider a sequence of n iid generalized Bernoulli trials with r possible outcomes with probabilities $p_1, \ldots p_r$.

Define $\vec{X} = (X_1, X_2, \ldots X_r)$, $X_i = \#$ of outcomes of type i.

Recall that if $\vec{n} = (n_1, n_2, \ldots n_r)$ $0 \le n_i \le n$
$\sum n_i = 1$

then

$$p_{\vec{X}}(\vec{n}) = \binom{n}{n_1, n_2, \ldots n_r} p_1^{n_1} p_2^{n_2} \ldots p_r^{n_r}$$

What is your intuition for the marginal distribution of X_i? Hint: remake the generalized Bernoulli trial as a standard Bernoulli trial.

Consider $r=3$. What is the probability that $X_1 = m$?

$$\sum_{k=0}^{n-m} \binom{n}{m, k, n-m-k} p_1^m p_2^k p_3^{n-m-k} =$$

$$p_1^m \sum_{k=0}^{n-m} \frac{n!}{m!(n-m)!} \frac{(n-m)!}{k!(n-m-k)!} p_2^k p_3^{n-m-k} =$$

$$\binom{n}{m} p_1^m \sum_{k=0}^{n-m} \binom{n-m}{k} p_2^k p_3^{n-m-k}$$

$$= \binom{n}{m} p_1^m (p_2+p_3)^{n-m}$$

"Remember" $(a+b)^N = \sum_{k=0}^{N} \binom{N}{k} a^k b^{N-k}$,

the binomial theorem.

p. 104 #1

Independent Random Variables

Discrete RV's X and Y are independent iff

$$P_{(X,Y)}(x,y) = P_X(x) P_Y(y)$$

Consequence

X, Y independent $\Rightarrow P(X \in A \cap Y \in B)$

$= P(X \in A) P(Y \in B)$

Verify: $\sum_{x \in A} \sum_{y \in B} P_{(X,Y)}(x,y) =$

$$\sum_{x \in A} \sum_{y \in B} P_X(x) P_Y(y) = \sum_{x \in A} P_X(x) \left(\sum_{y \in B} P_Y(y) \right)$$

$$= \sum_{x \in A} P_X(x) P(Y \in B) = P(Y \in B) \sum_{x \in A} P_X(x)$$

$$= P(X \in A) P(Y \in B)$$

Note $P(X=x/Y=y) = \dfrac{p_{XY}(x,y)}{p_Y(y)} = \dfrac{p_X(x)\cancel{p_Y(y)}}{\cancel{p_Y(y)}}$

$= p_X(x)$.

Def $X_1, \ldots X_r$ are independent if

$$p_{\vec{X}}(x_1, \ldots x_r) = p_{X_1}(x_1)\, p_{X_2}(x_2) \cdots p_{X_r}(x_r)$$

Note that this is stronger than pairwise independence.

Sum of Independent Random Variables:
If X and Y are independent random variables and $Z = X + Y$

$$P(Z=t) = \sum_{x=0}^{t} p_X(x)\, p_Y(t-x) = p_Z(t)$$

$\left(= \sum_{x=0}^{t} P(X = x \text{ \& } Y = t-x \right.$: what does this mean for $x \in \{x_1, x_2, \ldots x_s \ldots\}$?)

Theorem: If $X_1, X_2, \ldots X_r$ are mutually independent RV's with non negative integer images and $Z = X_1 + X_2 + \ldots X_r$ then

$$G_Z(z) = \prod_{i=1}^{r} G_{X_i}(z)$$

The proof in the case $r=2$ follows:

$$Z = X + Y$$

$$G_Z(z) = \sum_{t=0}^{\infty} p_Z(t) z^t$$

$$= \sum_{t=0}^{\infty} z^t \left(\sum_{x=0}^{t} p_X(x) p_Y(t-x) \right)$$

$$= \sum_{t=0}^{\infty} \sum_{x=0}^{t} p_X(x) z^x p_Y(t-x) z^{t-x}$$

$$= \sum_{x=0}^{\infty} \sum_{t=x}^{\infty} p_X(x) z^x p_Y(t-x) z^{t-x}$$

$$= \sum_{x=0}^{\infty} p_X(x) z^x \left(\sum_{t=x}^{\infty} p_Y(t-x) z^{t-x} \right)$$

set $y = t-x$

$$= \sum_{x=0}^{\infty} p_X(x) z^x \left(\sum_{y=0}^{\infty} p_Y(y) z^y \right)$$

$$= \sum_{x=0}^{\infty} p_X(x) z^x \, G_Y(z)$$

$$= G_X(z) G_Y(z).$$

The proof for the general case uses induction on r.

This, together with knowledge of the forms of the PGF's of the basic discrete random variables, enables us to identify the distributions of certain sums.

a. If $X_1, X_2, \ldots X_r$ are binomial with parameters n_i and p, then

$$X = \sum_{i=1}^{r} X_i \text{ is binomial with parameters}$$

$$n = \sum_{i=1}^{r} n_i \text{ and } p.$$

Intuition?

b. If X_i has the (modified) neg. binomial distribution with parameters α_i and p then

$$\sum_{i=1}^{r} X_i \quad \text{has the (modified)}$$

neg. binomial distribution with parameters

$\sum_{i=1}^{r} \alpha_i$ and p.

Intuition?

c. If X_i has the Poisson distribution with parameter α_i then $\sum_{i=1}^{r} X_i$ has the Poisson distribution with parameter $\sum_{i=1}^{r} \alpha_i$

Pf.

$G_{X_i}(z) = e^{-\alpha_i(1-z)}$, computed earlier.

Thus $G_Z(z) = \prod_{i=1}^{r} e^{-\alpha_i(1-z)}$

$$= e^{\sum_{i=1}^{r} -\alpha_i(1-z)}$$

$$= e^{(1-z)\sum_{i=1}^{r} -\alpha_i}$$

$$= e^{-\left(\sum_{i=1}^{r}\alpha_i\right)(1-z)},$$

the PGF for a Poisson distribution with parameter $\sum_{i=1}^{r}\alpha_i$

Note that a (modified) geometric distribution is a (modified) negative binomial distribution with $\alpha = 1$

Returning to the question of the finish time for concurrent processes, one with completion time X and one with completion time Y note the completion time for the process is (assuming X & Y independent)

$$Z = \max(X, Y).$$

$$F_Z(t) = P(Z \le t) = P(X \le t \text{ \& } Y \le t)$$

$$= F_X(t) F_Y(t)$$

If the completion time of the process is the minimum of X and Y,

$$Z = \min\{X, Y\}$$

$$P(Z > t) = P(X > t \;\&\; Y > t)$$

assuming X, Y ind. $= (1 - F_X(t))(1 - F_Y(t))$

$$= 1 - F_X(t) - F_Y(t) + F_X(t)F_Y(t)$$

so $F_Z(t) = F_X(t) + F_Y(t) - F_X(t)F_Y(t)$

Example: If X and Y have modified geometric distributions with parameter p, X, Y ind.

$$p_X(i) = p(1-p)^i \quad p_Y(j) = p(1-p)^j$$

so $F_X(k) = \sum_{i=0}^{k} p(1-p)^i = p \dfrac{1-(1-p)^{k+1}}{1-(1-p)}$

$= 1 - (1-p)^{k+1}$, likewise for F_Y.

Thus if $Z = \min\{X, Y\}$

$$F_Z(k) = 2(1-(1-p)^{k+1}) - (1-(1-p)^{k+1})^2$$

$$= 2 - 2(1-p)^{k+1} - (1 - 2(1-p)^{k+1} + (1-p)^{2(k+1)})$$

$$= 1 - ((1-p)^2)^{k+1}$$

Thus Z also has a modified geometric distribution with parameter p_3 satisfying

$$(1-p_3) = (1-p)^2$$

solving : $p_3 = 2p - p^2$

Example: Suppose X and Y are modified geometric with parameter p, so $X+Y$ is modified negative binomial with parameters 2 and p, then

$$P(Y=y \mid X+Y=t) =$$

$$\frac{P(Y=y \;\&\; X+Y=t)}{P(X+Y=t)}$$

$$= \frac{P(X=t-y)P(Y=y)}{P(X+Y=t)}$$

$$= \frac{\left(p(1-p)^{t-y}\right)\left(p(1-p)^{y}\right)}{\binom{t+2-1}{2-1} p^2 (1-p)^t} = \frac{(1!)(t!)}{(t+1)!}$$

$$= \frac{1}{t+1}$$, uniformly distributed over $\{0, 1, \dots t$

Problem 1 p. 112