MTH 7241 Fall 2022: Prof. C. King

Notes 1: Basics of probability

A number X generated by a random experiment is called a **random variable**. To understand X you must know its **range** $\mathrm{Ran}(X)$ (the set of possible values for X) and its **pdf** (the probabilities for those values).

Discrete Random Variables

If X is *discrete* then $\mathrm{Ran}(X)$ is finite or countably infinite, and the pdf p_X is the list of probabilities for each value. The pdf of a discrete r.v. X is a list of probabilities $\{p_X(x)\}$ where the index x runs over all possible values of X. If the range of X is small then this list can be written down explicitly. If the range of X is large or infinite then the pdf is given by a formula. The normalization condition says that the total probability for X must be one, that is

$$\sum_x p_X(x) = 1$$

where the sum runs over all possible values of X.

Example 1 (Bernoulli) *A Bernoulli random variable takes only two values, so it is determined by a single probability. The possible values are usually labeled* $\{0, 1\}$ *so the pmf is*

$$p_X(0) = \mathbb{P}(X = 0) = 1 - p, \quad p_X(1) = \mathbb{P}(X = 1) = p.$$

Example 2 *Roll of a fair die, sum of two rolls, maximum of two rolls.*

Example 3 (Binomial) *We write $X \sim \mathrm{Bin}(n,p)$ to indicate that X is a binomial r.v. where n is the number of independent trials and p is the probability of success on each trial. The formula for the pdf is*

$$p_X(k) = \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n$$

The standard example comes from coin tossing. Suppose a coin is tossed n times, and p is the probability of Heads on each toss (if $p = 1/2$ the coin is fair, otherwise it is biased). Let X be the number of times the coin comes up Heads.

HHHH, THHH, HTHT, TTHT,
HHHT, HHTT, THHT, THTT,
HHTH, TTHH, HTTH, HTTT,
HTHH, THTH, TTTH, TTTT

Then $X \sim \mathrm{Bin}(n,p)$. The explicit formula for p_X is useful if n and k are not too large. For example, an airline knows that 5% of people will not show up for a flight, so they overbook 52 people on a plane with 50 seats. What is the probability that nobody is bumped off the flight?

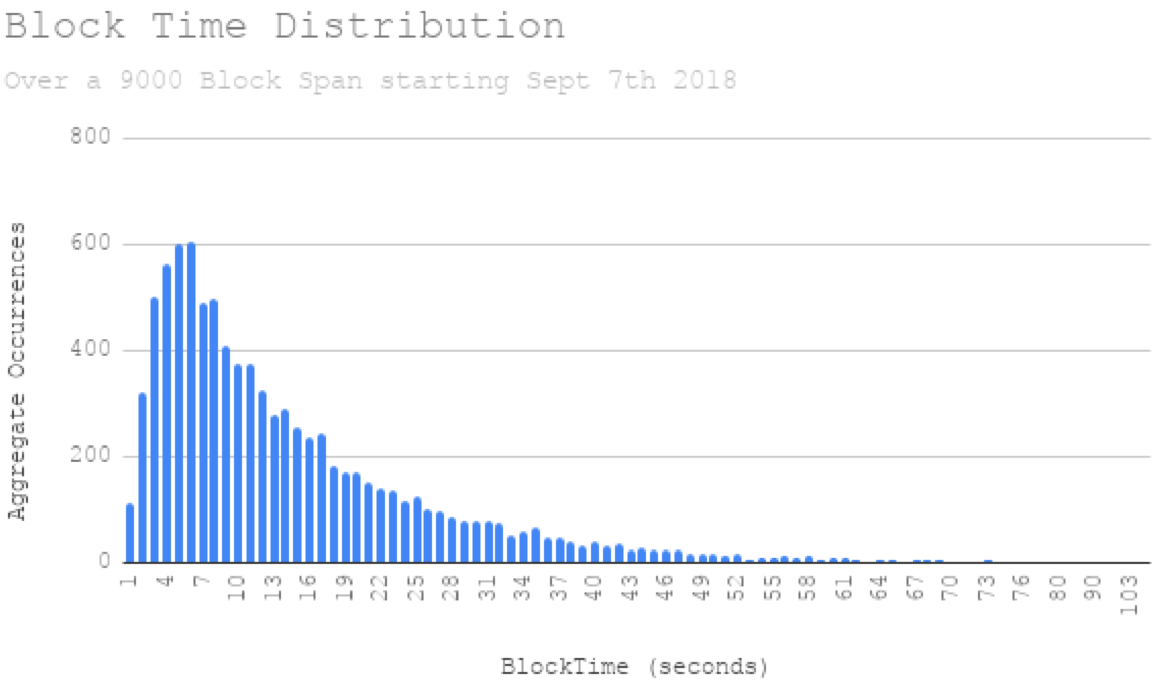
Example 4 (Poisson) *A Poisson random variable has infinite range* $0, 1, 2, \ldots$. *The pdf is given by the formula*

$$P(X = k) = \frac{\lambda^k}{k!}\, e^{-\lambda}, \quad k = 0, 1, \ldots$$

where $\lambda > 0$ *is a fixed parameter. Note that the Poisson is a useful approximation for the binomial* $\mathrm{Bin}(n, p)$ *(and is much simpler) in the case where* n *is large and* p *is small. The binomial convergence result is*

$$\lim_{\substack{n\to\infty \\ p\to 0 \\ np=\lambda}} \binom{n}{k}\, p^k\, (1-p)^{n-k} = \frac{\lambda^k}{k!}\, e^{-\lambda}$$

Compare Poisson value with exact result for previous example.


Block Time Distribution
Over a 9000 Block Span starting Sept 7th 2018
Aggregate Occurrences
800
600
400
200
0
1
4
7
10
13
16
19
22
25
28
31
34
37
40
43
46
49
52
55
58
61
64
67
70
73
76
80
90
103
BlockTime (seconds)

Continuous Random Variables

If X is *continuous* then Ran(X) is an interval of real numbers, and the pdf is the density of probabilities around each possible value. The pdf for a continuous r.v. X is a non-negative function f_X called the probability density function. Roughly, the value $f_X(x)\,dx$ is the probability to measure X in the interval $[x, x+dx]$ where dx is a small value. More precisely, the probability to find X in the interval $[a,b]$ is

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x)\,dx$$

Note that $\mathbb{P}(X = a) = \int_a^a f_X(x)\,dx = 0$, so the probability to find any particular value of X is zero. The normalization condition is

$$\int_{-\infty}^{\infty} f_X(x)\,dx = 1$$

Many special cases are important, we list a few here.

Example 5 (Uniform) *The r.v. X is* uniform *on the interval $[a,b]$ if X is 'equally likely' to be anywhere in the interval $[a,b]$, and has zero probability to be outside this interval. The pdf of X is constant on the interval $[a,b]$ and is zero outside this interval. That is,*

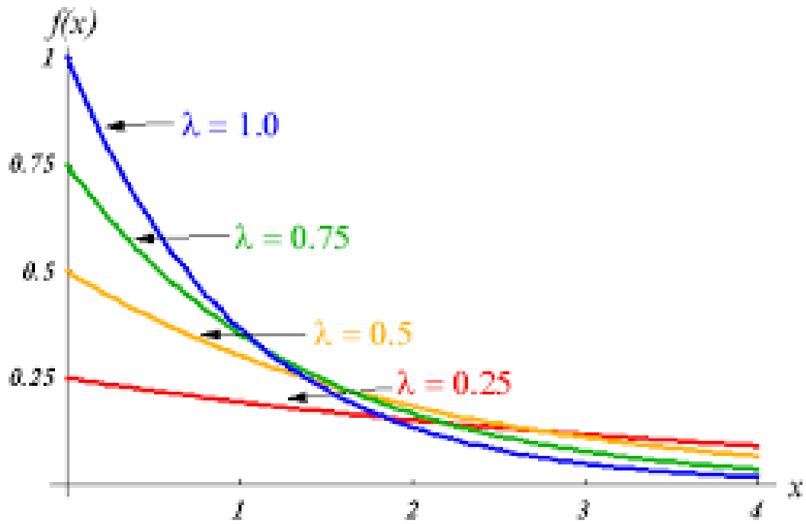
$$f_X(x) = \begin{cases} \frac{1}{b-a} & \textit{for } a \leq x \leq b \\ 0 & \textit{otherwise} \end{cases}.$$

The special case $[a,b] = [0,1]$ is often denoted by U, for uniform.

Example 6 (Exponential) *The pdf is*

$$f(x) = \begin{cases} ke^{-kx} & \textit{for } x \geq 0 \\ 0 & \textit{for } x < 0 \end{cases}$$

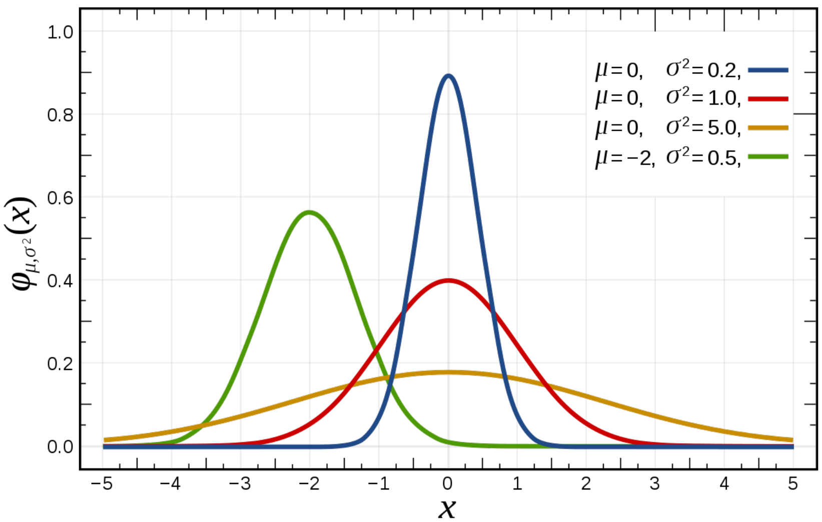
where $k > 0$ is called the rate. *This is often used to model the time until failure of a device.*



Example 7 (Normal) *Probably the most important family of r.v.'s, widely used in statistics. The pdf is*

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

where μ is the mean and $\sigma^2 > 0$ is the variance. The special case $\mu = 0$, $\sigma = 1$ is called the standard normal.



We finish up with a summary of properties of the most well-known (and useful) random variables. The first table lists the random variables along with their parameters and a useful 'story' that describes how they arise.

Name	Story
Bernoulli(p)	Toss a coin with probability p of turning up heads. X = Number of heads in one toss.
Binomial(n, p)	Toss a coin with probability p of turning up heads. X= Number of heads in n tosses. Binomial(n, p) is the sum of n independent Bernoulli(p).
Geometric(p)	Toss a coin with probability p of turning up heads. X = Number of tosses until the first Head.
Poisson(λ)	Random calls arrive with rate λ. X = Number of calls that arrive in one time unit.
Exponential(λ)	Random calls arrive with rate λ. X = time until the first arrival.
Gamma(n, λ)	Random calls arrive with rate λ. X = time until the nth arrival.
Uniform(a, b)	Pick a random number X between a and b.
Normal(μ, σ^2)	Pick an individual in a large population. X = height of the individual.

The second table lists the means, variances and pdf's of these random variables.

Name	Mean	Variance	pdf
Bernoulli(p)	p	$p(1-p)$	$p_X(k) = p^k\,(1-p)^{1-k}$
Binomial(n,p)	np	$np(1-p)$	$p_X(k) = \begin{pmatrix} n \\ k \end{pmatrix} p^k(1-p)^{n-k}$
Geometric(p)	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$p_X(k) = (1-p)^{k-1}p$
Poisson(λ)	λ	λ	$p_X(k) = e^{-\lambda}\lambda^k/k!$
Exponential(λ)	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$f_X(x) = \lambda e^{-\lambda x}$
Gamma(n,λ)	$\frac{n}{\lambda}$	$\frac{n}{\lambda^2}$	$f_X(x) = \lambda^k x^{k-1} e^{-\lambda x}/(k-1)!$
Uniform(a,b)	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$f_X(x) = 1/(b-a)$
Normal(μ,σ^2)	μ	σ^2	$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\,e^{-(x-\mu)^2/2\sigma^2}$

For a continuous random variable X, we can recover the pdf from the probability formula by noting that

$$f_X(x) = \frac{d}{dx} \int_{-\infty}^{x} f_X(u)\, du = \frac{d}{dx}\, \mathbb{P}(X \le x)$$

The *cumulative distribution function = cdf* is defined by

$$F_X(x) = \mathbb{P}(X \le x)$$

So we have the important relation

$$f_X(x) = \frac{d}{dx}\, F_X(x)$$

The cdf is important because it is often easier to first compute the cdf of a continuous r.v. and then take its derivative to get the pdf. Properties of the cdf:

(a) $0 \le F(x) \le 1$ for all $x \in \mathbb{R}$

(b) if $x < y$ then $F(x) \le F(y)$

(c) $\lim_{x \to \infty} F(x) = 1$, $\lim_{x \to -\infty} F(x) = 0$

(d) F is right continuous: if x_n is a decreasing sequence and $\lim x_n = x$ then $\lim F(x_n) = F(x)$

(e) $P(a < X \le b) = F(b) - F(a)$

(f) $P(X = x) = F(x) - \lim_{h \downarrow 0} F(x - h)$

Example 8 *Suppose U is uniform on $[0,1]$. Find the pdf of $X = \sqrt{U}$. Let's first find the cdf of X:* $\mathrm{Ran}(X) = [0,1]$*, and for any* $x \in [0,1]$

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(\sqrt{U} \leq x) = \mathbb{P}(U \leq x^2) = x^2$$

Therefore

$$f_X(x) = F_X(x)' = \begin{cases} 2x & \textit{for } 0 \leq x \leq 1 \\ 0 & \textit{else} \end{cases}$$

Example 9 *This example illustrates how 'randomness' can be a resource for computation. Suppose we want to find the area below the curve $y = f(x)$ over the range $0 \leq x \leq 1$, where*

$$f(x) = \left(3 - \sin^2 x\right)^{-1/2}$$

The integral cannot be computed exactly so we need some numerical method. Let U, V be independent uniform random numbers on $[0, 1]$. Think of these as the (x, y) coordinates of a randomly chosen point in the unit square. Then

$$\mathbb{P}\left(V \leq f(U)\right) = \textit{area under the curve } \{y = f(x)\} = \int_0^1 \left(3 - \sin^2 x\right)^{-1/2} \, dx$$

This provides our strategy to estimate the integral: generate N independent pairs (U_i, V_i), $i = 1, \ldots, N$, and compute the fraction

$$R_N = \frac{\#\{i : V_i \leq f(U_i)\}}{N}$$

As $N \to \infty$ this will converge to $\mathbb{P}(V \leq f(U))$. Here are results for different values of N, computed using Matlab:

N	*Trial 1*	*Trial 2*	*Trial 3*	*Time per trial (sec)*
1000	*0.617*	*0.595*	*0.590*	*0.0005*
10000	*0.6153*	*0.6154*	*0.6061*	*0.003*
100000	*0.6066*	*0.6079*	*0.6071*	*0.02*
1000000	*0.6074*	*0.6067*	*0.6073*	*0.18*

The actual value of the integral is 0.6071 *so with $N = 10^6$ we are getting a good estimate.*

Joint pdf

For two discrete random variables X and Y, the *joint pdf* is the list of probabilities for all possible pairs of values:

$$p_{X,Y}(x, y) = P(X = x, Y = y)$$

for all $x \in \text{Ran}(X)$ and $y \in \text{Ran}(Y)$.

If $\text{Ran}(X)$ and $\text{Ran}(Y)$ are finite, then the joint pdf is conveniently presented as a table of values.

Example 10 (Joint pdf: example) *Roll two dice, let X be the value on the first die, and let Y be the maximum of the two values.*

	X=1	*2*	*3*	*4*	*5*	*6*
Y=1	$\frac{1}{36}$	*0*	*0*	*0*	*0*	*0*
2	$\frac{1}{36}$	$\frac{2}{36}$	*0*	*0*	*0*	*0*
3	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{3}{36}$	*0*	*0*	*0*
4	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{4}{36}$	*0*	*0*
5	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{5}{36}$	*0*
6	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{6}{36}$

For continuous random variables X and Y, the joint pdf is the function $f_{X,Y}(x,y)$ of the two real variables x, y. Again the rough meaning is that $f_{X,Y}(x,y)\,dx\,dy$ is the probability to find the pair (X,Y) in the small rectangle $[x, x+dx] \times [y, y+dy]$. The precise meaning is that the probability for the pair (X,Y) to be in some region R in the xy-plane is the double integral

$$\mathbb{P}((X,Y) \in R) = \int_R f_{X,Y}(x,y)\,dx\,dy$$

We will not be much concerned with joint pdf's for continuous random variables in this course.

Joint pdf: marginals

We can recover the individual pdf of X from the joint pdf of (X, Y) by summing over all values of Y. Similarly we can recover the pdf of Y by summing over X.

Example 11 *Using our example above where X is the value on the first die, and Y is the maximum of the two values. Then we get p_X by holding the column fixed and summing over the rows, and we get p_Y by holding the row fixed and summing over the columns.*

	X=1	*2*	*3*	*4*	*5*	*6*
Prob	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

	Y=1	*2*	*3*	*4*	*5*	*6*
Prob	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

Expected value

The *expected value or mean* of a discrete random variable X is defined to be

$$\mathbb{E}[X] = \sum_i x_i \, P(X = x_i)$$

where the sum runs over the possible values of X.

The operational meaning is that $\mathbb{E}[X]$ is the long-run average value of repeated measurements of the random variable X. That is, suppose that we measure the random variable X in N independent trials, and record the results as $X_1, X_2, \dots, X_N$. Then the long-run average value is

$$\overline{X}_N = N^{-1} \, (X_1 + X_2 + \cdots + X_N)$$

We will shortly see the Law of Large Numbers which implies that

$$\lim_{N \to \infty} \overline{X}_N = \mathbb{E}[X]$$

For example, if X is the outcome of rolling a die, then the numbers $\{1, 2, \dots, 6\}$ all occur with probability $1/6$. So the expected value is

$$\mathbb{E}[X] = 1(1/6) + 2(1/6) + \cdots + 6(1/6) = 7/2$$

Example 12 *Roll three fair dice. Let X be the number of different faces that appear. Find the pdf of X and compute $\mathbb{E}[X]$.*

Example 13 *A biased coin has probability p of coming up Heads. Toss the coin until Heads first appears, let N be the number of tosses needed. The pdf of N is a geometric distribution. Compute $\mathbb{E}[N]$.*

Example 14 *A function of a random variable: $Y = g(X)$. Then*

$$P(Y = y) = \sum_{k:y=g(x_k)} P(X = x_k)$$

The expected value of Y can be computed like this:

$$\begin{aligned}
\mathbb{E}[Y] &= \sum_y y\, P(Y = y) \\
&= \sum_y \sum_{k:g(x_k)=y} g(x_k)\, P(X = x_k) \\
&= \sum_y \sum_k 1_{g(x_k)=y}\, g(x_k)\, P(X = x_k) \\
&= \sum_k g(x_k) P(X = x_k) \sum_y 1_{g(x_k)=y} \\
&= \sum_k g(x_k) P(X = x_k)
\end{aligned}$$

So the result is

$$\mathbb{E}[g(X)] = \sum_k g(x_k) P(X = x_k)$$

One very important property of the expected value is linearity, meaning that for any random variables X, Y and any real numbers a, b,

$$\mathbb{E}[aX + bY] = a\,\mathbb{E}[X] + b\,\mathbb{E}[Y]$$

This extends to a sum over any finite collection of r.v.'s:

$$\mathbb{E}[a_1X_1 + \cdots + a_nX_n] = a_1\mathbb{E}[X_1] + \cdots + a_n\mathbb{E}[X_n]$$

Variance

The mean $\mathbb{E}[X]$ is called the first order statistic of X. The second order statistic is the *variance*, defined by

$$\mathrm{VAR}[X] = \mathbb{E}\left[X - \mathbb{E}[X]\right]^2 = \sum_i (x_i - \mathbb{E}[X])^2 \, P(X = x_i)$$

Some elementary algebra yields

$$\mathrm{VAR}[X] = \mathbb{E}[X^2] - \left(\mathbb{E}[X]\right)^2$$

The standard deviation is defined as the square root of the variance:

$$\mathrm{STD}[X] = \sqrt{\mathrm{VAR}[X]}$$

Example 15 *Geometric: toss biased coin until first appearance of Heads, N is number of tosses needed. Saw that*

$$\mathbb{E}[N] = \frac{1}{p}$$

where p is the bias. Compute

$$\mathbb{E}[N(N-1)] = \frac{2(1-p)}{p^2}$$

and deduce $\mathrm{VAR}[N]$.

Joint pdf: Second order statistics

In addition to the means $\mathbb{E}[X]$ and $\mathbb{E}[Y]$ and the variances, there is another statistic which measures the relation between X and Y: the *Covariance* is

$$\mathrm{COV}[X,Y] = \mathbb{E}(X-\mathbb{E}[X])(Y-\mathbb{E}[Y]) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

The units depend on X and Y, so convenient to define the dimensionless *correlation coefficient*:

$$\mathrm{CORR}[X,Y] = \frac{\mathrm{COV}[X,Y]}{\mathrm{STD}[X]\,\mathrm{STD}[Y]}$$

Can show that $-1 \leq \mathrm{CORR}[X,Y] \leq 1$.

Joint pdf: Independence

Recall that two events A, B are *independent* if

$$\mathbb{P}(A \,\text{and}\, B) = \mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$$

The discrete random variables X and Y are *independent* if the events $\{X = x_i\}$ and $\{Y = y_j\}$ are independent for every pair of possible values (x_i, y_j). Equivalently, the joint pdf is the product of the marginal pmf's:

$$P(X = x_i, Y = y_j) = P(X = x_i)\,P(Y = y_j) \quad \text{for all } (x_i, y_j)$$

This definition extends immediately to the continuous case: the two continuous r.v.'s X and Y are independent if the joint pdf is the product of the marginals, that is if

$$f_{X,Y}(x, Y = f_X(x)\,f_Y(y) \quad \text{for all } x, y$$

Similarly, the r.v.'s $X_1, X_2, \ldots, X_n$ are independent if

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = P(X_1 = x_1)\,P(X_2 = x_2)\ldots P(X_n = x_n)$$

for all possible values $x_1, \ldots, x_n$.

Example 16 *Coin tosses are independent. Suppose we toss a coin twice and*

$$X = \begin{cases} 1 & \textit{if first toss is Heads} \\ 0 & \textit{first toss is Tails} \end{cases}$$
$$Y = \begin{cases} 1 & \textit{if second toss is Heads} \\ 0 & \textit{second toss is Tails} \end{cases}$$

Suppose the coin is biased and p *is the probability of Heads, so* $1-p$ *is the probability of Tails. Then for example*

$$\mathbb{P}(X = 1, Y = 0) = \mathbb{P}(X = 1)\,\mathbb{P}(Y = 0) = p(1 - p)$$

Example 17 *Consider again our example above where X is the value on the first die, and Y is the maximum of the two values. It's easy to check that X and Y are dependent. [For example, there cannot be a row or column with some entries zero and others non-zero].*

Joint pdf: Mean of an independent product

We noted that the expected value is a *linear operator* on random variables:

$$\begin{aligned}\mathbb{E}[X+Y] &= \mathbb{E}[X]+\mathbb{E}[Y] \\ \mathbb{E}[aX+bY] &= a\,\mathbb{E}[X]+b\,\mathbb{E}[Y]\end{aligned}$$

However the expected value of a product is generally *not* the product of expected values, except in one important case: *if* X, Y *are independent* then

$$\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$$

This can be seen as follows:

$$\begin{aligned}\mathbb{E}[XY] &= \sum_{i,j} x_i y_j \, P(X=x_i, Y=y_j) \\ &= \sum_{i,j} x_i y_j \, P(X=x_i)P(Y=y_j) \\ &= \sum_i x_i P(X=x_i) \sum_j y_j P(Y=y_j) \\ &= \mathbb{E}[X]\,\mathbb{E}[Y]\end{aligned}$$

Similarly for any functions of independent random variables we get

$$\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\,\mathbb{E}[h(Y)]$$

[Note that if X and Y are independent then $\mathrm{CORR}[X,Y]=0$. In general the converse is false, but in many cases it is 'mostly true' ...]

Joint pdf: Variance of an independent sum

As a consequence, *if* X, Y *are independent* then

$$\mathrm{VAR}[X+Y] = \mathrm{VAR}[X] + \mathrm{VAR}[Y]$$

To see this, note that

$$\begin{aligned}
\mathrm{VAR}[X+Y] &= \mathbb{E}[X+Y]^2 - (\mathbb{E}[X] + \mathbb{E}[Y])^2 \\
&= \mathbb{E}[X^2] + \mathbb{E}[Y^2] + 2\mathbb{E}[X]\mathbb{E}[Y] \\
&\quad -(\mathbb{E}[X])^2 - (\mathbb{E}[Y])^2 - 2\mathbb{E}[X]\mathbb{E}[Y] \\
&= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 + \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2
\end{aligned}$$

This same result applies to a sum of many independent random variables. If $X_1, \ldots, X_n$ are *independent* then

$$\mathrm{VAR}[X_1 + \cdots + X_n] = \mathrm{VAR}[X_1] + \cdots + \mathrm{VAR}[X_n]$$

Example 18 *Binomial r.v. X can be written as a sum of Bernoulli r.v.'s. If $X \sim Bin(n,p)$ then*

$$X = X_1 + X_2 + \cdots + X_n$$

where each X_i is Bernoulli with probability $P(X_i = 1) = p$. Then it easily follows that

$$\mathbb{E}[X] = np, \quad \mathrm{VAR}[X] = np(1-p)$$

Example 19 *Coupon collecting problem: there are m types of coupons. Every cereal box has one randomly selected coupon. Let N be the number of boxes collected until we get at least one coupon of every type. Question: what is $\mathbb{E}[N]$?*

As we collect boxes, we encounter new types. Let X_i be the number of boxes after encountering the $(i-1)^{st}$ new type until we encounter the i^{th} new type. Then

$$N = X_1 + X_2 + \cdots + X_m$$

and each X_i is geometric with probability of success

$$p_i = 1 - \left(\frac{i-1}{m}\right) = \frac{m-i+1}{m}$$

Then

$$\mathbb{E}[N] = \sum_{i=1}^{m} \frac{1}{p_i} \simeq m\,\log m$$

Example 20 *An urn contains n Red balls and m Black balls. Suppose that k balls are withdrawn from the urn, and let X be the number of Red balls among these. Find $\mathbb{E}[X]$ assuming (i) replacement, and (ii) no replacement.*

The key to solving this problem is to rewrite X as a sum of many small random variables which can each be easily analyzed, and then use the distributive properties of the expected value to put these together for X. For each $i = 1, \ldots, k$ define

$$R_i = \begin{cases} 1 & \textit{if the } i\textit{th ball is Red} \\ 0 & \textit{if the } i\textit{th ball is Black} \end{cases}$$

Then it should be clear that

$$X = R_1 + R_2 + \cdots + R_k$$

Consider R_1: its pdf is

R_1	0	1
Probability	$\frac{m}{n+m}$	$\frac{n}{n+m}$

So we compute

$$\mathbb{E}[R_1] = \frac{n}{n+m}$$

What about $R_2, R_3, \ldots$? Assuming replacement, it is clear that these all have the same pdf's (and they are independent), so we also have

$$\mathbb{E}[R_i] = \frac{n}{n+m} \quad \textit{for all } i = 1, \ldots, k$$

and this implies

$$\mathbb{E}[X] = \frac{k\,n}{n+m}$$

Suppose we do not replace the balls after drawing. Then the variables $R_1, R_2, \ldots$ are not independent, however they all have the same marginal pdf. To see this, consider R_2:

$$\mathbb{P}(R_2 = 1) = \mathbb{P}(R_1 = 0, R_2 = 1) + \mathbb{P}(R_1 = 1, R_2 = 1)$$

Now we have

$$\mathbb{P}(R_1 = 0, R_2 = 1) = \mathbb{P}(R_1 = 1, R_2 = 0)$$

(To see this, think about drawing the first two balls at the same time, and suppose one is Red and the other is Black. Which ball was drawn first? It doesn't matter, so both orders are equally likely). Then we have

$$\begin{aligned} \mathbb{P}(R_2 = 1) &= \mathbb{P}(R_1 = 1, R_2 = 0) + \mathbb{P}(R_1 = 1, R_2 = 1) \\ &= \mathbb{P}(R_1 = 1) \end{aligned}$$

This means that R_1 and R_2 have the same pdf, so in particular

$$\mathbb{E}[R_2] = \mathbb{E}[R_1] = \frac{n}{n+m}$$

The same argument can be extended to show that if $k \leq n + m$ then

$$\mathbb{E}[R_i] = \frac{n}{n+m} \quad \textit{for all } i = 1, \ldots, k$$

and so $\mathbb{E}[X]$ is the same as in the case with replacement. For the homework problem you will analyze the variance of X, which does turn out to be different for the two cases!