Math 425
Intro to Probability
Lecture 27

Kenneth Harris
kaharri@umich.edu

Department of Mathematics
University of Michigan

March 20, 2009

Problem involving sums

Problem. Let X and Y be two independent discrete random variables with known probability mass functions p_X and p_Y and whose possible values are only nonnegative integers.

What is the probability mass function for $X + Y$?

☞ Consider a possible event $\{X + Y = n\}$. Then

$$\begin{aligned} p_{X+Y}(n) &= \mathbf{P}\{X+Y=n\} = \mathbf{P}\{Y = n - X\} \\ &= \sum_{k=0}^{n} \mathbf{P}\{X=k,\ Y=n-k\} \\ &= \sum_{k=0}^{n} \mathbf{P}\{X=k\} \cdot \mathbf{P}\{Y=n-k\} \\ &= \sum_{k=0}^{n} p_X(k) \cdot p_Y(n-k). \end{aligned}$$

Convolutions: discrete case

Definition

Let X and Y be discrete random variables taking only nonnegative values.

The convolution of X and Y is the probability mass function $p = p_X * p_Y$ given by

$$p(n) = \sum_{k=0}^{n} p_X(k) \cdot p_Y(n-k).$$

$p_X * p_Y$ is the probability mass function for the sum $X + Y$.

Note. Generalizing the convolution to discrete random variables with integer or rational possible values is straightforward. However, the cases we are interested in (Poisson, Geometric, Binomial, Uniform) fit this definition.

Example

Example. A die is rolled twice. Let X and Y be the outcomes, so they have the common mass function

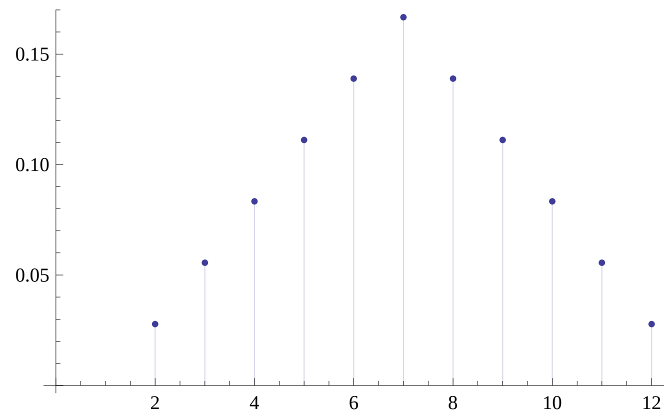
$$p(i) = \begin{cases} \frac{1}{6} & \text{if } 1 \le i \le 6 \\ 0 & \text{othewise} \end{cases}$$

☞ The convolution $p_{X+Y} = p_X * p_Y$ is computed as follows

$$\begin{aligned} p(2) &= p_1(1) \cdot p_2(1) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36} \\ p(3) &= p_1(1) \cdot p_2(2) + p_1(2) \cdot p_2(1) = \frac{2}{36} \\ p(4) &= p_1(1) \cdot p_2(3) + p_1(2) \cdot p_2(2) + p_1(3) \cdot p_2(1) = \frac{3}{36} \\ p(n) &= \begin{cases} \frac{1}{36}(n-1) & \text{if } 2 \le n \le 7 \\ \frac{1}{36}(13-n) & \text{if } 8 \le n \le 12 \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

Example

☞ Sum of two uniform discrete random variables on $\{1, 2, 3, 4, 5, 6\}$.



Key Properties

☞ The convolution has several important properties which are especially useful for computing the probability mass function of the sum of several independent random variables.

Theorem

Let X, Y and Z be discrete random variables taking only nonnegative values.

(a) *(Commutativity).* $p_X * p_Y = p_Y * p_X$.

(b) *(Associativity).* $(p_X * p_Y) * p_Z = p_X * (p_Y * p_Z)$.

Sums of random variables

☞ Let $X_1, X_2, \ldots, X_n$, be discrete random variables which are independent and identically distributed with distribution $p_X(x)$.

Consider the random variable which sums these

$$S_n = \sum_{k=1}^{n} X_k \qquad n \geq 1.$$

whose distribution is

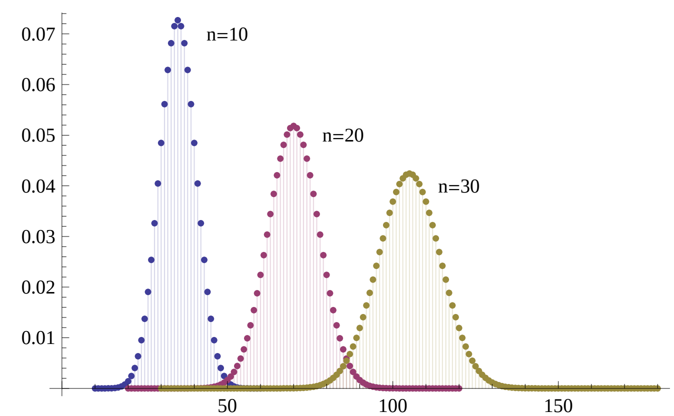
$$\begin{aligned} p_{S_n} &= p_{X_1} * p_{X_2} * \cdots * p_{X_n} \\ &= p_{S_{n-1}} * p_{X_n} \\ &= p_{S_{n-1}} * p_X. \end{aligned}$$

It is possible in some cases to calculate the distribution for S_n by recursion from the distribution for S_{n-1}.

Example

☞ Sum of $n = 10, 20, 30$ uniform discrete random variables on $\{1, 2, 3, 4, 5, 6\}$. It is a bit tedious to compute ☺

Looks more and more like a Bell curve. As n increases the distribution approaches the normal distribution with $\mu = n \cdot 3.5$ and $\sigma^2 = n \cdot \frac{35}{12}$.



Theorem

☞ A Bernoulli random variable gives the number of successes in one Bernoulli trial:

$$p(0) = 1 - p \qquad p(1) = p.$$

☞ A binomial random variable gives the number of successes in n Bernoulli trials:

$$p(k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad 0 \le k \le n.$$

Theorem (Binomial Random Variables)

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed Bernoulli random variables with parameter $p \in (0,1)$.

The sum S_n of these random variables is a binomial random variable with pararmeters n and p.

Proof

Basis. S_1 is a single Bernoulli random variable, so a binomial random variable on 1 trial.

Inductive Step. Suppose that the sum S_n of n independent and identically distributed Bernoulli random variables with probability $p \in (0,1)$ has the distribution

$$p_{S_n}(k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad 0 \le k \le n.$$

☞ Let X_{n+1} be a Bernoulli random variable independent of those in S_n and with the same distribution. Show $p_{S_{n+1}} = p_{S_n} * p_{X_{n+1}}$ has distribution

$$p_{S_{n+1}}(k) = \binom{n+1}{k} p^k (1-p)^{n+1-k} \qquad 0 \le k \le n+1.$$

Proof – continued

☞ Compute $S_{n+1} = S_n + X_{n+1}$ using $p_{S_n} * p_{X_{n+1}}$.

$$\begin{aligned}
\mathbf{P}\{S_n + X_{n+1} = k\} &= \sum_{j=0}^{k} p_{S_n}(k-j) \cdot p_{X_{n+1}}(j) \\
&= p_{S_n}(k) \cdot (1-p) + p_{S_n}(k-1) \cdot p \\
&= \binom{n}{k} p^k (1-p)^{n+1-k} + \binom{n}{k-1} p^k (1-p)^{n+1-k} \\
&= \binom{n+1}{k} p^k (1-p)^{n+1-k} \qquad \boxtimes
\end{aligned}$$

where $0 \le k \le n+1$.

☞ I owe you ✉, but this show that the sum of $n+1$ independent Bernoulli random variables is a binomial random variable on $n+1$ trials.

Proof – continued

✉ Justify (when $k \le n$)

$$\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1}.$$

☞ How many ways are there of choosing a set of size k from $\{1, 2, \ldots, n+1\}$?

1. When $n+1$ IS NOT in the subset, then choose the k members from $\{1, 2, \ldots, n\}$, so

$$\binom{n}{k} \qquad \text{possibilities.}$$

2. When $n+1$ IS in the subset, then choose the other $k-1$ members from $\{1, 2, \ldots, n\}$, so

$$\binom{n}{k-1} \qquad \text{possibilities.}$$

Theorem

☞ A geometric random variable gives the waiting time for the first success

$$p(m) = p(1-p)^{m-1} \qquad m = 1, 2, 3, \ldots$$

☞ A negative binomial random variable with parameter n gives the waiting time for the nth success.

$$p(m) = \binom{m-1}{n-1} p^n (1-p)^{m-n} \qquad m = n, n+1, n+2, \ldots$$

Theorem (Geometric Random Variables)

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed geometric random variables with probability $p \in (0,1)$.

The sum S_n of these random variables is a negative binomial random variable with parameters n and p.

Proof

Basis. S_1 is a single geometric random variable, so a negative binomial random variable with $n = 1$.

Inductive Step. Suppose that the sum S_n of n independent and identically distributed geometric random variables with probability $p \in (0,1)$ has the distribution

$$p_{S_n}(m) = \binom{m-1}{n-1} p^n (1-p)^{m-n} \qquad m = n, n+1, n+2, \ldots$$

☞ Let X_{n+1} be a geometric random variable independent of those in S_n and with the same distribution. Show $p_{S_{n+1}} = p_{S_n} * p_{X_{n+1}}$ has distribution

$$p_{S_{n+1}}(m) = \binom{m-1}{n} p^{n+1} (1-p)^{m-n-1} \qquad m = n+1, n+2, n+3, \ldots,$$

Proof – continued

☞ Let $m \geq n+1$ (the minimal possible value). Compute $p_{S_n} * p_{X_{n+1}}$

$$\begin{aligned}
\mathbf{P}\{S_n + X_{n+1} = m\} &= \sum_{j=0}^{m} p_{S_n}(j) \cdot p_{X_{n+1}}(m-j) \\
&= \sum_{j=n}^{m-1} \binom{j-1}{n-1} p^n (1-p)^{j-n} \cdot p(1-p)^{m-j-1} \\
&= \sum_{j=n}^{m-1} \binom{j-1}{n-1} p^{n+1} (1-p)^{m-n-1} \\
&= \binom{m-1}{n} p^m (1-p)^{n-m} \qquad ✉
\end{aligned}$$

☞ I owe you ✉, but this show that the sum of $n+1$ independent geometric random variables is a negative binomial random variable with parameters $n+1$ and p.

Proof – continued

✉ Justify (when $m \geq n$)

$$\binom{m}{n} = \sum_{j=n}^{m} \binom{j-1}{n-1}$$

☞ How many ways are there of choosing a set of size n from $\{1, 2, \ldots, m\}$?

1. Choose the largest possible value in $\{n, n+1, \ldots, m\}$.
2. For each $j = n, \ldots m$ (the largest value in the subset of size n), choose a subset of size $n-1$ from $\{1, \ldots, j-1\}$. There are

$$\binom{j-1}{n-1} \qquad \text{possibilities.}$$

Definition: Convolution

☞ How are sums of independent random variables distributed?

Analogous to the definition for discrete random variables, we define the convolution of continuous random varariables.

Definition

Let X and Y be two continuous random variables with densities $f_X(x)$ and $f_Y(y)$.

The convolution $f = f_X * f_Y$ is the function given by

$$f(z) = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y)\, dy$$

Note. The convolution is commutative and associative: for any density functions f, g and h,

$$\begin{aligned} f * g &= g * f \\ (f*g)*h &= f*(g*h). \end{aligned}$$

Theorem: convolutions and sums of r.v.'s

Theorem

Let X and Y be independent continuous random variables with density $f_X(x)$ and $f_Y(y)$.

The sum $X+Y$ is a continuous random variable with density $f_{X+Y} = f_X * f_Y$. That is

$$f_{X+Y}(a) = \int_{-\infty}^{\infty} f_X(a-y) f_Y(y)\, dy$$

Proof

Proof. Let X and Y be independent with joint density $f_{X,Y}(x,y)$.

$$\begin{aligned} F_{X+Y}(a) &= \mathbf{P}\{X+Y \le a\} = \mathbf{P}\{X \le a - Y\} \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{a-y} f_{X,Y}(x,y)\, dx\, dy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{a-y} f_X(x) f_Y(y)\, dx\, dy \\ &= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{a-y} f_X(x)\, dx\right] f_Y(y)\, dy \\ &= \int_{-\infty}^{\infty} F_X(a-y) f_Y(y)\, dy. \end{aligned}$$

Proof – continued

$$F_{X+Y}(a) = \int_{-\infty}^{\infty} F_X(a-y) f_Y(y)\, dy.$$

☞ We get the density for $X+Y$ by differentiating F_{X+Y}

$$\begin{aligned} f_{X+Y}(a) &= \frac{d}{da} F_{X+Y}(a) \\ &= \frac{d}{da}\int_{-\infty}^{\infty} F_X(a-y) f_Y(y)\, dy \\ &= \int_{-\infty}^{\infty} \frac{d}{da} F_X(a-y) f_Y(y)\, dy \\ &= \int_{-\infty}^{\infty} f_X(a-y) f_Y(y)\, dy = f_X * f_Y(a) \end{aligned}$$

Theorem

☞ An exponential random variable gives the waiting time for the first success in a Poisson process

$$f(t) = \lambda e^{-\lambda t} \qquad t \geq 0$$

☞ A gamma random variable with parameters $(\alpha = n, \lambda)$ is the waiting time for the n success in a Poisson process

$$f(t) = \lambda e^{-\lambda t}\frac{(\lambda t)^{n-1}}{(n-1)!} \qquad t \geq 0.$$

Theorem (Exponential Random Variables)

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed exponential random variables with parameter $\lambda > 0$.

Then the sum S_n of these random variables is a gamma random variable with pararmeters $(\alpha = n, \lambda)$.

Proof

Basis. S_1 is a single exponential random variable, so a gamma random variable with $(\alpha = 1, \lambda)$.

Inductive Step. Suppose that the sum S_n of n independent and identically distributed exponential random variables with probability $p \in (0,1)$ has the distribution

$$f_{S_n}(t) = \lambda e^{-\lambda t}\frac{(\lambda t)^{n-1}}{(n-1)!} \qquad t \geq 0.$$

☞ Let X_{n+1} be an exponential random variable independent of those in S_n and with the same distribution. Show $p_{S_{n+1}} = p_{S_n} * p_{X_{n+1}}$ has distribution

$$f_{S_{n+1}}(t) = \lambda e^{-\lambda t}\frac{(\lambda t)^{n}}{(n)!} \qquad t \geq 0.$$

Proof – continued

☞ Compute

$$\begin{aligned} f_{S_n+X_{n+1}}(a) &= \int_{-\infty}^{\infty} f_{S_n}(y)\cdot f_{X_{n+1}}(a-y)\,dy \\ &= \int_0^a \lambda e^{-\lambda y}\frac{(\lambda y)^{n-1}}{(n-1)!}\cdot \lambda e^{-\lambda(a-y)}\,dy \\ &= \lambda e^{-\lambda a}\int_0^a \lambda\frac{(\lambda y)^{n-1}}{(n-1)!}\,dy \\ &= \lambda e^{-\lambda a}\frac{(\lambda a)^n}{n!} \end{aligned}$$

☞ This show that the sum of $n+1$ independent exponential random variables is a gamma random variable with parameters $(\alpha = n+1, \lambda)$.

Theorem

☞ The sum of independent gamma random variables is a gamma random variable.

Theorem (Gamma Random Variables)

Let $X_1, X_2, \ldots, X_n$ be independent gamma distributed random variables with parameters (α_i, λ), $i = 1, \ldots, n$.

Then their sum S_n is also a gamma distributed random variable but with parameter $(\sum_{i=1}^n \alpha_i, \lambda)$.

☞ See Ross Proposition 3.1 for the case of $n = 2$, but the proof is similar to the case of exponential random variables. When the α_i are integers, this theorem follows from the result about sums of exponential random variables.

Theorem

Theorem (Normal Random Variables)

Let $X_1, X_2, \ldots, X_n$ be independent normal random variables with respective parameters μ_i, σ_i^2, $i = 1, \ldots, n$.

Then the sum S_n of these random variables is a normal random variable with pararmeters $\sum_{i=1}^{n} \mu_i$ and $\sum_{i=1}^{n} \sigma_i^2$.

☞ See Ross Proposition 6.3.2, page 283.

Proof of a special case

☞ Ross proves the most general case. I will prove the case of two standard normal random variables, X and Y, whose density is

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

with mean $\mu = 0$ and variance $\sigma^2 = 1$.

☞ When X is a normally distributed random variable with mean μ and variance σ^2, the density is

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

Proof of a special case

Then the density of $X + Y$ is

$$\begin{aligned}
f_{X+Y}(a) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-(a-y)^2/2} e^{-y^2/2}\, dy \\
&= \frac{1}{2\pi} e^{-a^2/2} \int_{-\infty}^{\infty} e^{-(y^2 - ay)}\, dy \\
&\overset{1}{=} \frac{1}{2\pi} e^{-a^2/4} \int_{-\infty}^{\infty} e^{-(y-(a/2))^2}\, dy \\
&\overset{2}{=} \frac{1}{2\pi} e^{-a^2/4} \sqrt{\pi} \left[\frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-(y-(a/2))^2}\, dy \right] \\
&= \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-a^2/4} \qquad (\text{normal: } \mu = 0, \sigma^2 = 2)
\end{aligned}$$

(1): Complete the square: $-(y^2 - ay) = -(y^2 - ay + (a/2)^2) + (a/2)^2$
(2): $[\ldots] = 1$: normal with $\mu = \frac{a}{2}$ and $\sigma^2 = \frac{1}{2}$.

Proof of a special case

☞ When X and Y are normally distributed random variables with mean $\mu = 0$ and $\sigma^2 = 1$, their sum $X + Y$ has density

$$f_{X+Y}(a) = \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-a^2/4}$$

and is a normally distributed random variable with $\mu = 0$ and $\sigma^2 = 2$.