Lecture 1: Jan 5. Review of random variables Ross 4.2-4.5, 5.1-5.2

1.1 Definitions

(i) **Definition:** A random variable X is a real-valued function on the sample space.

(ii) **Definition:** A random variable X is *discrete* if it can take only a discrete set of values.

(iii) **Definition:** A *continuous* random variable X is one that takes values in $(-\infty, \infty)$. That is, in principle. In practice, some values may be impossible.

1.2 Examples

(i) Discrete (finite): the number of heads in 10 tosses of a fair coin.

(ii) Discrete (countable): the number of traffic accidents in a large city in a year.

(iii) Continuous (bounded range): A random number between a and b: values in the interval (a, b).

(iv) Continuous (unbounded range): The waiting time until the bus arrives: values in $(0, \infty)$.

1.3 Probability mass function (pmf) or density (pdf)

(i) **Definition:** The *probability mass function* (p.m.f.) of a discrete random variable X is the set of probabilities $P(X = x)$ for each of the values $x \in \mathcal{X}$ that X can take.

(ii) $P(X = x) \;\geq\; 0$ for each $x \in \mathcal{X}$ and $\sum_x P(X = x) \;=\; 1$ where the sum is over all $x \in \mathcal{X}$.

(iii) **Definition:** The *probability density function* (p.d.f.) of a continuous random variable X is a non-negative function f_X defined for all values x in $(-\infty, \infty)$ such that for any subset B for which $X \in B$ is an event

$$P(X \in B) \;=\; \int_B f_X(x)\; dx$$

(iv) X takes some value in $(-\infty, \infty)$ so

$$1 \;=\; P(-\infty < X < \infty) \;=\; \int_{-\infty}^{\infty} f_X(x)\; dx$$

1.4 Examples

(i) Example (i): *Binomial*: $B(10, \frac{1}{2})$. $\quad P(X = x) \;=\; \binom{10}{x}(1/2)^{10}$ for $x = 0, 1, 2, ...10$.

(ii) Example (ii): *Poisson*: $Po(\mu)$ mean μ. $\quad P(X = x) \;=\; \exp(-\mu)\mu^x/x!$ for $x \;=\; 0, 1, 2, 3, 4.....$

(iii) Example (iii): *Uniform p.d.f*:

$$f_X(x) \;=\; \frac{1}{(b-a)} \text{ for } a \leq x \leq b \text{ and } f_X(x) \;=\; 0 \text{ otherwise.}$$

(iv) Example (iv): *Exponential p.d.f*

$$f_X(x) \;=\; \lambda \exp(-\lambda x) \text{ for } x \geq 0 \text{ and } f_X(x) \;=\; 0 \text{ if } x < 0.$$

1.5 Expectations of (functions of) random variables

(i) Discrete case (Ross, 4.3):

If X is discrete with p.m.f. $P(X = x) \;=\; p_X(x) > 0$ for $x \in \mathcal{X}$, the *expected value* of X denoted $\mathrm{E}(X)$ is $\mathrm{E}(X) \;=\; \sum_{x \in \mathcal{X}} x\; p_X(x)$, provided this sum exists and is finite.

(ii) Continuous case (Ross 5.2)

If X is continuous with p.d.f. $f_X(x)$, $f_X(x) \geq 0$ for $-\infty < x < \infty$. the *expected value* of X denoted $\mathrm{E}(X)$ is $\mathrm{E}(X) \;=\; \int_{-\infty}^{\infty} x f_X(x)\; dx$, provided this integral exists and is finite. (Note $f_X(x)\; dx \;\approx\; P(x \;<\; X \;\leq\; x+dx)$.)

(iii) Functions of a random variable:

$$\mathrm{E}(g(X)) \;=\; \sum_x g(x) p_X(x) \text{ (discrete), or } \mathrm{E}(g(X)) \;=\; \int_x g(x) f_X(x)\; dx \text{ (continuous).}$$

(iv) Variance: If $E(X) = \mu$, $\mathrm{var}(X) \;=\; \mathrm{E}(X-\mu)^2$. In fact, $\mathrm{var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2$. Note $\mathrm{var}(X) \geq 0$.

Lecture 2: Jan 7. Review of Continuous random variables Ross 5.3-5.

2.1 The probability density function: definition and basic properties.

(i) For a subset of the real line B:

$$P(X \in B) \;=\; \int_B f_X(x)\,dx$$

(ii) In fact, events can be made up of unions and intersections of intervals of the form $(a, b]$:

$$P(a < X \leq b) \;=\; \int_a^b f_X(x)\,dx$$

(iii) Note the value at the boundary does not matter:

$$P(X = a) \;=\; \int_a^a f_X(x)\,dx \;=\; 0 \quad \text{for any continuous random variable.}$$

(iv) Note: $f_X(x) = 0$ is possible for some x-values (see the p.m.f).
For example, if $X \geq 0$ (as in the waiting-time example), $f(x) = 0$, if $x < 0$.

2.2 The *cumulative distribution function* of X is

$$F_X(x) \;=\; P(-\infty < X \leq x)$$

The cdf is defined for any random variable, but it is most useful for continuous random variables. In this case

$$F_X(x) \;=\; \int_{-\infty}^{x} f_X(z)\,dz \quad \text{and} \quad f_X(x) = \frac{d}{dx}F_X(x)$$

2.3 The Uniform distribution on (a, b).

$$f(x) \;=\; \frac{1}{(b-a)} \text{ for } a \leq x \leq b \text{ and } f(x) \;=\; 0 \text{ otherwise.}$$

If $a = 0$ and $b = 1$, $f_X(x) = 1$ and $F_X(x) = x$ on $0 < x < 1$. $\mathrm{E}(X) = 1/2$, $\mathrm{var}(X) = 1/12$.

2.4 The exponential distribution with rate parameter λ: $\mathcal{E}(\lambda)$.

$$f(x) \;=\; \lambda\exp(-\lambda x) \text{ for } x \geq 0 \text{ and } f(x) \;=\; 0 \text{ if } x < 0.$$

$F_X(x) \;=\; 1 - \exp(-\lambda x)$ for $x > 0$. $\mathrm{E}(X) = 1/\lambda$, $\mathrm{var}(X) = 1/\lambda^2$.

2.5 The Normal distribution with mean μ **and variance** σ^2

$$f_X(x) \;=\; \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \qquad -\infty < x < \infty$$

$\mathrm{E}(X) = \mu$. $\mathrm{var}(X) = \sigma^2$.
If $X \sim N(\mu, \sigma^2)$, then $(X - \mu)/\sigma \;\sim\; N(0, 1)$.

2.6 Location and scale

(i) A location parameter a shifts a probability density: the pdf is a function of $(x - a)$. For example, we can shift a uniform $U(0, 1)$ pdf to a uniform $U(a, a + 1)$ pdf. If $X \sim U(0, 1)$, $Y = a + X \sim U(a, a + 1)$.

(ii) A scale parameter stretches (or shrinks) a probability density. For example, to transform a Uniform $U(0, 1)$ density to a Uniform $U(a, b)$, we shift by a and scale by $(b - a)$. If $X \sim U(0, 1)$, $Y = a + (b - a)X \sim U(a, b)$.

(iii) The parameter λ^{-1} of an exponential random variable is also a scale parameter.
If $X \sim \mathcal{E}(\lambda)$, then $\lambda X \sim \mathcal{E}(1)$. If $X \sim \mathcal{E}(\lambda)$, then $Y = kX \sim \mathcal{E}(\lambda/k)$.

(iv) A Normal random variable has both location μ and scale σ.
If $X \sim N(\mu, \sigma^2)$, then $(X - \mu)/\sigma \;\sim\; N(0, 1)$. If $X \sim N(\mu, \sigma^2)$, then $Y = aX + b \;\sim\; N(a\mu + b, a^2\sigma^2)$.

Lecture 3; Jan 9. Review of the Bernoulli process

3.1: The process

0 0 1 0 0 0 1 1 0 0 1 0 0 0 0 0 0 0 1 0 0 1 0 0 1: Each *trial* is *success* (1) or not (0).

$X_1 X_2 X_3 X_4 ....$ $...X_{25}...$: Each X_i is 0 or 1.

... ... T_5... ... ... T_{10}... ... ... T_{15}... ... ... T_{20}... ... ... T_{25}...: $T_n = X_1 + ... + X_n$

$---Y_1 ----Y_2 Y_3 --Y_4 ----------Y_5 --Y_6 ---Y_7$: Y_r is rth inter-arrival time

... ...W_1... ... ...$W_2 W_3$.. ..W_4... ... ...W_5... ...W_6... ...W_7: W_r is total waiting time to rth 1.

The Bernoulli process is defined by X_i independent, with $P(X_i = 1) = p$ and $P(X_i = 0) = (1-p)$.

$T_n = X_1 + ... + X_n$ is number of *successes* (i.e. 1s) in first n trials.

Y_r is the *inter-arrival time*: number of trials from $(r-1)$th success to r th.

$W_r = Y_1 + Y_2 + ... + Y_r$ is number of trials to r th. success.

$Y_r^* = Y_r - 1$: number of failures (0) before next success (1).

$W_r^* = Y_1^* + ... + Y_r^*$: number of failures (0) before r th success. **Note:** $W_r > n$ if and only if $T_n < r$.

3.2: Bernoulli and Binomial random variables Ross 4.6

(i) X_i is Bernoulli(p). $P(X_i = 1) = p$, $P(X_i = 0) = (1-p)$.

$\mathrm{E}(X_i) = p \times 1 + (1-p) \times 0 = p$

$\mathrm{E}(X_i^2) = p \times 1^2 + (1-p) \times 0^2 = p$, so $\mathrm{var}(X_i) = \mathrm{E}(X_i^2) - (\mathrm{E}(X_i))^2 = p - p^2 = p(1-p)$.

$T_n = X_1 + ... + X_n$ is Binomial (n, p).

The probability of each sequence of k 1's and $(n-k)$ 0's is $p^k(1-p)^{n-k}$ and there are $\binom{n}{k}$ such sequences.

$P(T_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$.

Expectations always add: $\mathrm{E}(T_n) = \mathrm{E}(X_1) + ... + \mathrm{E}(X_n) = p + p + ... + p = np$.

In general variances do NOT add, but here they do: $\mathrm{var}(T_n) = \mathrm{var}(X_1) + ... + \mathrm{var}(X_n) = np(1-p)$.

3.3: Geometric and Negative Binomial random variables Ross 4.8.1, 4.8.2

Y_r are independent, and have Geometric (p) distribution: $P(Y = k) = (1-p)^{k-1}p$, for $k = 1, 2, 3, ...$

$\mathrm{E}(Y) = \sum_{k=1}^{\infty} k(1-p)^{k-1}p = p/(1-(1-p))^2 = 1/p$. $\mathrm{var}(Y) = (1-p)/p^2$.

$Y^* = (Y-1)$, $P(Y^* = k) = (1-p)^k p$, for $k = 0, 1, 2, 3....$

$\mathrm{E}(Y_r^*) = \mathrm{E}(Y) - 1 = (1-p)/p$, $\mathrm{var}(Y^*) = \mathrm{var}(Y)$. (Recall $\mathrm{E}(aY + b) = a\mathrm{E}(Y) + b$ and $\mathrm{var}(aY + b) = a^2 \mathrm{var}(Y)$).

$W_r = Y_1 + ... + Y_r$. Expectations add, so $\mathrm{E}(W_r) = r/p$. Again the variances do add, $\mathrm{var}(W_r) = r(1-p)/p^2$.

$P(W_r = k) = P(r-1 \text{ successes in } k-1 \text{ trials, and then success}) = \binom{k-1}{r-1}(1-p)^{k-r}p^{r-1}p$ for $k = r, r+1, ...$

$W_r^* = W_r - r$, $\mathrm{E}(W_r^*) = \mathrm{E}(W_r) - r$, $\mathrm{var}(W_r^*) = \mathrm{var}(W_r)$

$P(W_r^* = k) = P(r-1 \text{ successes in } r+k-1 \text{ trials, and then success}) = \binom{r+k-1}{r-1}(1-p)^k p^{r-1} p$

for $k = 0, 1, 2, 3...$

3.4 A reminder about the hypergeometric distribution Ross 4.8.3

Example: the number of red fish, in sampling n fish without replacement, from a pond in which there are N fish of which m are red.

$$P(X = x) = \binom{m}{x}\binom{N-m}{n-x} / \binom{N}{n} \text{ for } x = \max(0, m+n-N), ... \min(m, n).$$

Lecture 4; Jan 12. Introduction to the Poisson process Ross 4.7, 9.1

4.1 The process

Events occur *randomly and independently* in time, at rate λ.

More formally: the numbers of events N in disjoint time intervals are independent, and the probability distribution of the number of events $N(\ell)$ in an interval depends only on its length, ℓ. Additionally, $P(N(h) = 1) = \lambda h + o(h)$, $P(N(h) \geq 2) = o(h)$.

4.2 The waiting time T to an event

The waiting time T to an event is $> s$, if there are no events in $(0, s)$. That is $P(T > s) \; = \; P(N(s) = 0) \; \equiv \; P_0(s)$.

$$
\begin{aligned}
P_0(s+h) \; &= \; P_0(s) \times P_0(h) \; = \; P_0(s)(1 - \lambda h - o(h)) \\
P_0(s+h) - P_0(s) \; &= \; -\lambda h P_0(s) + o(h) \\
dP_0/P_0 \; &= \; -\lambda ds \quad \text{or} \quad log(P_0) = -\lambda s \;\; \text{with} \;\; P_0(0) = 1 \\
\text{So} \quad P(T > s) \; &= \; P_0(s) \; = \; \exp(-\lambda s) \\
\text{So} \qquad F_T(s) \; &= \; P(T \leq s) \; = \; 1 - P(T > s) \; = \; 1 - \exp(-\lambda s) \\
\text{So} \qquad f_T(s) \; &= \; F_T'(s) \; = \; \lambda \exp(-\lambda s) \quad \text{on} \quad 0 < s < \infty
\end{aligned}
$$

That is, regardless of where we start waiting, the waiting time to an event is exponential with rate parameter λ. Recall the "forgetting property" of the exponential: $P(T > t + s | T > t) \; = \; P(T > s)$.

4.3 The number of events $N(s)$ in a time interval length s

Let $N(s)$ be the number of events in interval $(0, s)$ and $P_n(s) = P(N(s) = n)$.

Note from 4.2, $P_0(s) = P(T > s) = \exp(-\lambda s)$. Then

$$
\begin{aligned}
P_n(s+h) \; &= \; P_n(s)(1 - \lambda h - o(h)) + P_{n-1}(s)(\lambda h + o(h)) \; + \; o(h) \\
P_n(s+h) - P_n(s) \; &= \; \lambda h (P_{n-1}(s) - P_n(s)) \; + \; o(h) \\
P_n'(s) \; &= \; \lambda (P_{n-1}(s) - P_n(s)) \text{ letting } h \to 0
\end{aligned}
$$

Hence from $P_0(s) = \exp(-\lambda s)$ we could determine P_1, P_2, ...

Instead, consider $q_n(s) = \exp(-\lambda s)(\lambda s)^n / n!$.

Then $q_n'(s) \; = \; \exp(-\lambda s)\lambda^n n s^{n-1}/n! \; - \; \lambda \exp(-\lambda s)(\lambda s)^n/n! \; = \; \lambda(q_{n-1}(s) - q_n(s))$.

That is, $P_n(s) \equiv q_n(s)$. That is $N(s)$ is a Poisson random variable with mean λs.

4.4 The conditional distribution of times of events

Suppose we know exactly 1 event occurred in $(0, s)$. At what time T did it occur?

This is a continuous random variable: $P(T = t) = 0$ for every t.

Instead consider the cdf $P(T \leq t)$:

$$
\begin{aligned}
F_T(t) \; &= \; P(T \leq t \mid N(s) = 1) \; = \; P(T \leq t \; \cap \; N(s) = 1)/P(N(s) = 1) \\
&= \; P_1(t) P_0(s-t)/P_1(s) \; = \; (\lambda t \exp(-\lambda t)) \exp(-\lambda(s-t))/(\lambda s \exp(-\lambda s)) \; = \; t/s
\end{aligned}
$$

So $F_T(t) = t/s$ on $0 < t < s$, or $f_T(t) = 1/s$, $0 < t < s$.

That is T is uniform on the interval $(0, s)$.