Lecture 1: Random Variables and Convergence.

Zhongjian Wang*

Abstract

Classnotes on random variables, random number generation, distribution and convergence

1 Basic Notion and Examples

Consider throwing a die, there are 6 possible outcomes, denoted by ω_i, $i = 1, \cdots, 6$; the set of all outcomes $\Omega = \{\omega_1, \cdots, \omega_6\}$, is called *sample space*.

A subset of Ω, e.g. $A = \{\omega_2, \omega_4, \omega_6\}$, is called an *event*. Suppose we did N times of die experiment, event A happened N_a times, then the *probability* of event A is $P(A) = \lim_{N\to\infty} N_a/N$. For a fair die, $P(A) = 1/2$.

Let the collection of events be A, A a sigma-algebra of all events, meaning (1) if $E \in A$, then $E^c \in A$; (2) if $E_i \in A$, i countable, then $\cup_i E_i \in A$. The triple (Ω, A, P) is called a *probibility space*. P is a function assigning probability to events, more precisely, a probability measure satisfyng: (1) $P(E) \geq P(\Phi) = 0$, Φ null event, (2) if E_i are countably many disjoint events, $P(\cup_i E_i) = \sum_i P(E_i)$, (3) $P(\Omega) = 1$.

The events E and F are *independent*, if:

$$P(E \cap F) = P(E)P(F),$$

and *conditional probability* $P(E|F)$ is:

$$P(E|F) = P(E \cap F)/P(F).$$

A *random variable r.v.* $X(\omega)$ is a function: $\Omega \to R$ such that $\{\omega \in \Omega : X(\omega) \leq a\}$ is an event. The *distribution function* of $X(\omega)$ is:

$$F_X(x) = P(\{\omega \in \Omega : X(\omega) \leq x\}), \tag{1.1}$$

*Department of Statistics, University of Chicago

satisfying:
(1) $\lim_{x\to-\infty} F_X(x) = 0$, $\lim_{x\to+\infty} F_X(x) = 1$.
(2) $F_X(x)$ is nondecreasing, right continuous ($\{X \le y\} \to \{X \le x\}$ as $y \to x + 0$).
(3) $F_X(x-) = P(X < x)$ ($\{X \le y\} \to \{X < x\}$ as $y \to x - 0$).
(4) $P(X = x) = F_X(x) - F_X(x-)$.

Conversely, if F satisfies (1)-(3), it's a distribution function of some r.v.

When F_X is absolutely continuous, we have a density function $p(x)$ such that:

$$F(x) = \int_{-\infty}^{x} p(y)\,dy.$$

Examples (continuous r.v): (1) Uniform distribution on $[a, b]$:

$$p(x) = \chi_{[a,b]}(x)/(b - a);$$

(2) unit or standard Gaussian (normal) distribution:

$$p(x) = (2\pi)^{-1/2} e^{-x^2/2};$$

(3) exponential distribution ($\lambda > 0$):

$$p(x) = \lambda e^{-\lambda x} \chi_{(x\ge 0)};$$

Examples (discrete r.v): (1) two point r.v, taking x_1 with prob. p, x_2 with prob. $1 - p$, distribution is:

$$F_X = \begin{cases} 0 & x < x_1 \\ p & x \in [x_1, x_2) \\ 1 & x \ge x_2, \end{cases}$$

(2) Poisson distribution with ($\lambda > 0$):

$$p_n = P(X = n) = \lambda^n \exp\{-\lambda\}/n!,\ n = 0, 1, 2, \cdots.$$

Mean of a r.v. is:

$$\mu = E(X) = \sum_{j=1}^{N} x_j p_j,$$

the discrete case and:

$$\mu = E(X) = \int_{R^1} x p(x)\,dx,$$

the continuous case.

Variance is: $\sigma^2 = Var(X) = E((X - \mu)^2)$.

2 Random Number Generators

On digital computers, psuedo-random numbers are used as approximations of random numbers. A common algorithm is the linear recursive scheme:

$$X_{n+1} = aX_n \ \ (mod\, c\,), \tag{2.2}$$

a and c positive relatively prime integers, with initial value "seed" X_0. The numbers:

$$U_n = X_n/c,$$

will be approximately uniformly distributed over $[0,1]$. c is usually a large integer in powers of 2, a is a large integer relative prime to c.

Matlab command "rand(m,n)" generates $m \times n$ matrices with psuedo random entries uniformly distributed on $(0,1)$ ($c = 2^{1492}$), using current state. S = rand('state') is a 35-element vector containing the current state of the uniform generator. rand('state',0) resets the generator to its initial state. rand('state',J), for integer J, resets the generator to its J-th state. Similarly, "randn(m,n)" generates $m \times n$ matrices with psuedo random entries standard-normally distributed, or unit Gaussian.
Example: a way to visualize the generated random numbers is:

$$\begin{aligned}
&t = (0 : 0.01 : 1)';\\
&rand('state', 0);\\
&y1 = rand(size(t));\\
&randn('state', 0);\\
&y2 = randn(size(t));\\
&plot(t, y1,' b', t, y2,' g')
\end{aligned}$$

Two-point r.v. can be generated from uniformly distributed r.v. $U \in [0,1]$ as:

$$X = \begin{cases} x_1 & U \in [0,p] \\ x_2 & U \in (p,1] \end{cases}$$

A continuous r.v with distribution function F_X, can be generated from U as $X = F_X^{-1}(U)$ if F_X^{-1} exists, or more generally:

$$X = \inf\{x : \ U \le F_X(x)\}.$$

This is called inverse transform method. It applies to exponential distribution, to give:

$$X = -\ln(1-U)/\lambda,\ U \in (0,1).$$

The Box-Muller method generates Gaussian from two independent $U_i \in [0,1]$, $i = 1,2$ by a nonlinear mapping:

$$N_1 = \sqrt{-2\ \ln U_1}\ \cos(2\pi U_2),$$
$$N_2 = \sqrt{-2\ \ln U_1}\ \sin(2\pi U_2), \tag{2.3}$$

where N_1, N_2 are independent unit Gaussian.

Recall: two Gaussian distribution are independent if their covariance is 0.

3 Moment Inequalities

Some useful inequalities involving moments are:
• Markov inequality:

$$P(\{\omega : X(\omega) \geq a\}) \leq \frac{1}{a} E(X), \text{ if } X(\omega) \geq 0;$$

and Chebyshev inequality:

$$P(\{\omega : |X(\omega)|^2 \geq a\}) \leq \frac{1}{a} E(X^2),$$

for any $a > 0$.
• Jensen's inequality:

$$g(E(X)) \leq E(g(X)),\ \ g \text{ convex}.$$

It follows that for any $0 < r < s$:

$$(E(|X-a|^r))^{1/r} \leq (E(|X-a|^s))^{1/s},$$

Lyapunov inequality.
• Hölder inequality:

$$E(|X+Y|^r)^{1/r} \leq E(|X|^r)^{1/r} + E(|Y|^r)^{1/r},\ r \geq 1,$$

$$E(|X \cdot Y|) \leq (E(|X|^p))^{1/p}(E(|Y|^q))^{1/q},\ p^{-1} + q^{-1} = 1.$$

4 Joint Distribution

For n r.v's $X_1, X_2, \cdots, X_n$, *Joint Distribution Function* is:

$$F_{X_1,\cdots,X_n}(x_1,\cdots,x_n) = P(\{\omega \in \Omega : X_i(\omega) \le x_i, i = 1,2,\cdots,n\}).$$

• $n = 2$,

$$F_{X_1,X_2} \to 0,\ x_i \to -\infty,$$

$$F_{X_1,X_2} \to 1,\ x_1, x_2 \to +\infty,$$

F_{X1,X_2} is nondecreasing and right continuous in x_1 and x_2.
Marginal Distribution F_{X_1}:

$$F_{X_1}(x_1) = \lim_{x_2 \to \infty} F_{X_1,X_2}(x_1, x_2).$$

Continuous r.v:

$$F_{X_1,X_2}(x_1,x_2) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} p(y_1,y_2)dy_1dy_2,$$

$p \ge 0$ density.
Joint Gaussian with mean $\mu = (\mu_1, \mu_2)$ *and covariance* $C^{-1} = (E(X_i - \mu_i)(X_j - \mu_j)) > 0$:

$$p(x_1,x_2) = \frac{\sqrt{det(C)}}{2\pi} \exp\{\left(-\frac{1}{2}\sum_{i,j=1}^{2} c^{i,j}(x_i - \mu_i)(x_j - \mu_j)\right)\}, \tag{4.4}$$

orthogonal transformation of Gaussian r.v. is Gaussian.
• Independence:

$$F_{X_1X_2}(x_1,x_2) = F_{X_1}(x_1)F_{X_2}(x_2),$$

$$p(x_1,x_2) = p_1(x_1)p_2(x_2).$$

5 Convergence and Limit Theorems

Sequence of r.v. $X_1, X_2, \cdots, X_n$:

• convergence with prob 1 (wp1) , also called almost surely convergence (a.s.):

$$P\left(\{\omega \in \Omega : \lim_{n\to\infty} |X_n(\omega) - X(\omega)| = 0\}\right) = 1; \tag{5.5}$$

• mean-square convergence, also called L^2 convergence : $(E(X_i^2) \le C)$

$$\lim_{n\to\infty} E(|X_n - X|^2) = 0; \tag{5.6}$$

• convergence in probability (prob.):

$$\lim_{n\to\infty} P(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| \ge \epsilon\}) = 0, \; \forall\, \epsilon; \tag{5.7}$$

• convergence in law, convergence in distribution (d.):

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x), \tag{5.8}$$

at all continuous points of F_X;

• weak convergence (w):

$$\lim_{n\to\infty} \int_{R^1} f(x) dF_X(x) = \int_{R^1} f(x) dF_X(x), \tag{5.9}$$

for any $f \in C_0(R^1)$.

Following convergence Theorem holds:

1. a.s. $\Longrightarrow$ prob.
2. $L^2 \Longrightarrow$ prob.
3. DCT: If $|X_i| \le |Y|$, wp1, $E(|Y|^2) < \infty$, a.s. $\Longrightarrow L^2$
4. prob. $\Longrightarrow$ d. $\Longrightarrow$ w., for any $f \in C_0(R^1)$.

Example 1: i.i.d. r.v X_i's, with $\mu = E(X_i)$, $\sigma^2 = Var(X_i)$,

$$\frac{S_n}{n} = \frac{\sum_{i=1}^n X_i}{n} \to \mu,$$

a.s. and L^2 (Strong Law of Large Numbers), prob. (Weak Law of Large Numbers).

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \to_{d.} N(0,1),$$

,$N(0,1)$ unit Gaussian, Central Limit Theorem.

Example 2: Let $\Omega = [0,1]$, $P([a,b]) = |b-a|$, $[a,b] \subset [0,1]$. Let $A_n = \{\omega : \omega \in [0,1/n]\}, X_n = \sqrt{n}\chi_{A_n}$, then $X_n \to 0$ in probability, a.s, but not in L^2.

6 Project I (due: before lecture 4)

You can use any programming language, you don't need to tex your report. Please send the digital version to honglizhaobob@uchicago.edu . Title will be P1-(Your Name on UID).

I1. Generate $N = 10^4$ uniformly distributed pseudo random numbers on $(0, 1)$ on Matlab (or in other environment). Partition the interval into subintervals I_j of equal length 0.05. Count the number of random numbers in I_j as N_j. Plot relative frequencies N_j/N divided by subinterval length, so called histogram. Does the histogram look like density of $U(0, 1)$? Compute sample average:

$$\mu_N = \frac{1}{N}\sum_{j=1}^{N} x_j,$$

and sample variance:

$$\sigma_N = \frac{1}{N-1}\sum_{j=1}^{N}(x_j - \mu_N)^2.$$

Compare them to 1/2 and 1/12, exact mean and variance of $U(0, 1)$.

I2. Repeat I1 for random variable defined on $[0, 2]$ interval whose PDF is given by $\rho(x) = \frac{1}{4}x^3$.

I3. Show that the two random variables N_1, N_2 generated by Box-Muller method are Gaussian with zero mean and identity covariance when U_1, U_2 are independent $U(0, 1)$ uniformly distributed.

I4. (1) Let $Z = (N_1, N_2)$, S an invertible 2 x 2 matrix, $\mu \in R^2$, show that $X = S^T Z + \mu$ is jointly Gaussian with mean μ, and covariance matrix $S^T S$.

(2) Write a program to generate a pair of Gaussian pseudo random numbers (X_1, X_2) with zero mean and covariance $E(X_1^2) = 1$, $E(X_2^2) = 1/3$, $E(X_1 X_2) = 1/2$. Generate 1000 pairs of such numbers, evaluate their sample averages and sample covariances.

(3) Is it possible to generate a pair of real random variables (X_1, X_2) with zero mean and covariance $E(X_1^2) = 1$, $E(X_2^2) = 1/3$, $E(X_1 X_2) = 1$? Note the random variables does not have to be Gaussian.

You can get Matlab under UChicago license from the underlining website https://www.mathworks.com/academia/tah-portal/university-of-chicago-719588.html
I suggest you to install full package or at least include the statistics toolbox when installing. Again, you are free to use any programming language.