STAT 380 Lecture 25

Inverse (Probability Integral) Transform

Fact: F continuous CDF and X, U satisfy

$$U = F(X)$$

then $U \sim$Uniform(0,1) if and only if $X \sim F$.

Proof: For simplicity: assume F strictly increasing on inverval (a, b) with $F(b) = 1$ and $F(a) = 0$.

If $U \sim$Uniform(0,1) then

$$\begin{aligned} P(X \leq x) &= P(F(X) \leq F(x)) \\ &= P(U \leq F(x)) \\ &= F(x) \end{aligned}$$

Conversely: if $X \sim F$ and $0 < u < 1$ then there is a unique x such that $F(x) = u$

$$\begin{aligned} P(U \leq u) &= P(F(X) \leq F(x)) \\ &= P(X \leq x) \\ &= F(x) \\ &= u \end{aligned}$$

Application: generate U and solve $F(X) = U$ for X to get $X = F^{-1}(U)$ which has cdf F.

Example: For the exponential distribution

$$F(x) = 1 - e^{-\lambda x}$$

Set $F(X) = U$ and solve to get

$$X = -\log(1 - U)/\lambda$$

Observation: $U \sim 1 - U$ so $X = -\log(U)/\lambda$ also has an Exponential(λ) distribution.

Example: : for F the standard normal cdf solving

$$F(X) = U$$

requires numerical approximation but such solutions are built in to many languages.

Special purpose transformations:

Example: : If X and Y are iid $N(0, 1)$ define $R \geq 0$ and $\Theta \in [0, 2\pi)$ by

$$X = R\cos(\Theta) \qquad Y = R\sin(\Theta)$$

NOTE: Book says

$$\Theta = \tan^{-1}(Y/X)$$

but this takes values in $(-\pi/2, \pi/2)$; notation

$$\tan^{-1}(x, y)$$

means angle θ in $[0, 2\pi)$ so that

$$\frac{x}{\sqrt{x^2+y^2}} = \cos(\theta) \qquad \frac{y}{\sqrt{x^2+y^2}} = \sin(\theta)\,.$$

Then

$$\{(r, \theta) : r \le r^*, \theta \le \theta^*\}$$

is part of circle of radius r^*. (Start at $(r^*, 0)$ and go clockwise to angle θ^*.)

Compute joint cdf of R, Θ at r^*, θ^*.

Define

$$C = \{(x, y) : x^2 + y^2 \le (r^*)^2, \quad 0 \le \tan^{-1}(x, y) \le \theta^*\}$$

Then

$$P(R \le r, \Theta \le \theta) = \iint_C \frac{e^{-(x^2+y^2)/2}}{2\pi} dxdy$$

Do integral in polar co-ordinates to get

$$\int_0^{r^*} \int_0^{\theta^*} \frac{e^{-r^2/2}}{2\pi} r\, dr d\theta$$

The θ integral just gives

$$\theta^*/(2\pi)$$

The r integral then gives

$$1 - e^{-(r^*)^2/2}$$

So

$$P(R \le r^*, \Theta \le \theta^*) = \left(1 - e^{-(r^*)^2/2}\right) \times \frac{\theta^*}{2\pi}$$

Product of two cdfs so: R and Θ are independent.

Moreover R^2 has cdf

$$P(R^2 \le x) = 1 - e^{-x/2}$$

which is Exponential with rate 1/2.

Θ is Uniform$(0, 2\pi)$.

So: generate U_1, U_2 iid Uniform(0,1).

Define

$$\Theta = 2\pi U_2$$

and

$$R = \sqrt{-2\log(U_1)}$$

Put

$$X = R\cos(\Theta) \qquad Y = R\sin(\Theta)$$

You have generated two independent $N(0,1)$ variables.

Acceptance Rejection

If $F(X) = U$ difficult to solve (eg F hard to compute) sometimes use acceptance rejection.
Goal: generate $X \sim F$ where $F' = f$.
Tactic: find density g and constant c such that

$$f(x) \leq cg(x)$$

and s.t. can generate Y from density g.
Algorithm:

1. Generate Y which has density g.
2. Generate $U \sim$Uniform(0,1) independent of Y.
3. If $U \leq f(Y)/\{cg(Y)\}$ let $X = Y$.
4. If not reject Y and go back to step 1; repeat, generating new Y and U (independently) till you accept a Y.

Facts:

- Since you repeat the same experiment each trial: number of trials, N, success is Geometric (and sure to be finite).
- The probability of success on a individual trial is

$$\begin{aligned} &P[U \leq f(Y)/\{cg(Y)\}] \\ &\quad = \int P[U \leq f(y)/\{cg(y)\}|Y = y]g(y)dy \\ &\quad = \int \frac{f(y)}{cg(y)}g(y)dy \\ &\quad = \frac{1}{c}\int f(y)dy \\ &\quad = 1/c \end{aligned}$$

so

$$\mathrm{E}(N) = 1/c$$

Most important fact: $X \sim f$:

Proof: this is like the old craps example:

We compute

$$P(X \leq x) = P[Y \leq x | U \leq f(Y)/\{cg(Y)\}]$$

(condition on the first iteration where the condition is met).

Condition on Y to get

$$\begin{aligned} & P(X \leq x) \\ & = \frac{\int_{-\infty}^{x} P[U \leq f(Y)/\{cg(Y)\}|Y = y]g(y)dy}{P[U \leq f(Y)/\{cg(Y)\}]} \\ & = \frac{\int_{-\infty}^{x} \frac{f(y)}{cg(y)} g(y)dy}{1/c} \\ & = \int_{-\infty}^{x} f(y)dy \\ & = F(x) \end{aligned}$$

as desired.