Random Sample Generation & Extreme Values

Josh Engwer
Texas Tech University

February 18, 2012

GENERATING A RANDOM SAMPLE:

- **Probablity Integral Transform Theorem:**
  \( X \) is continuous r.v., \( F_X \) is strictly increasing cdf, \( U \sim \text{Uniform}(0, 1) \implies F_X \sim \text{Uniform}(0, 1) \) and \( F_X^{-1}(U) \sim X \)

- **REMARK:** The Probability Integral Transform Thm also holds for discrete r.v.’s provided \( F_X^{-1} \) is properly defined.

- **EXAMPLE:** \( X \sim \text{Exponential}(\lambda) \implies \text{cdf } F_X(x) = 1 - e^{-x/\lambda}, x > 0 \implies F_X^{-1}(u) = -\lambda \log (1 - u) \)
  
  Hence, to generate \( X_1, \ldots, X_n \sim \text{iid Exp}(\lambda) \), generate \( U_1, \ldots, U_n \sim \text{Uniform}(0, 1) \) and set each \( X_i = F_X^{-1}(U_i) \)

- **Definition of Generalized Inverse CDF:** general inverse of cdf \( F(\cdot) \) is \( F^{-} (u) := \inf \{ x : F(x) \geq u \} \)

- **Generalized Probability Integral Transform:**
  \( X \sim F_X \) and \( U \sim \text{Uniform}(0, 1) \implies F_X \sim \text{Uniform}(0, 1) \) and \( F_X^{-}(U) \sim X \)

- **Box-Muller Method for Generating Random Sample of Normal(0,1):**
  
  Generate \( U_1, U_2 \sim \text{iid Uniform}(0,1) \) r.v.’s and set \( \left\{ \begin{array}{l} R = \sqrt{-2 \log U_2} \\ \theta = 2\pi U_1 \end{array} \right\} \), then \( \left\{ \begin{array}{l} X = R \cos \theta \\ Y = R \sin \theta \end{array} \right\} \sim \text{iid Normal}(0,1) \)

- Other methods to generate random samples: Monte-Carlo, Accept/Reject, Metropolis Algorithm

**EXTREME VALUE (E-V) THEORY:** (Limit Theorems for the largest order statistic \( X_{(n)} \))

- Recall that cdf of \( X_{(n)} \) is \( F_{X_{(n)}}(x) = [F(x)]^n \) and \( X_{(1)} = \min \{ X_1, \ldots, X_n \} = -\max \{ -X_1, \ldots, -X_n \} \)

- **DEFINITION:** cdf \( F(\cdot) \) is **attracted to** cdf \( G(\cdot) \) (AKA \( F \) belongs to domain of attraction of \( G \))
  \[ X_1, \ldots, X_n \sim \text{iid } F(\cdot) \text{ and } \exists \text{ sequences } \{a_n\}, \{b_n\} \text{ s.t. } \frac{X_{(n)} - a_n}{b_n} \overset{d}{\to} Y \sim \text{ non-degenerate } G(\cdot) \]

- **FISHER-TIPPETT THEOREM:** \( F(\cdot) \) is attracted to \( G(\cdot) \implies G \) must be one of the following std E-V types:
  
  (i) Frechet : \( G_1(y; \gamma) = \text{Exp} \{ -1/y^\gamma \} I(y > 0), \text{ where } \gamma > 0 \)
  (ii) Weibull : \( G_2(y; \gamma) = \text{Exp} \{ (-y)^\gamma \} I(y < 0) + 1 \cdot I(y \geq 0), \text{ where } \gamma > 0 \)
  (iii) Gumbel : \( G_3(y) = \text{Exp} \{ -e^{-y} \} \)

- The above 3 extreme value types are related to \( E \sim \text{Exponential}(1) \) by : \( E^{-1/\gamma} \sim G_1, -E^{1/\gamma} \sim G_2, -\log E \sim G_3 \)

- Distributions which belong to domain of attraction of \( G_1 \): Cauchy, Pareto, ...

- Distributions which belong to domain of attraction of \( G_2 \): Beta, Uniform, ...

- Distributions which belong to domain of attraction of \( G_3 \): Exponential, Lognormal, Normal, Weibull, ...

- Distributions which have no domain of attraction: Discrete dist’s & continuous dist’s with heavy or wiggly tails.

---

Copyright 2012 Josh Engwer
References