15-750: Graduate Algorithms February 10, 2017

Lecture 11: Probability, Order Statistics and Sampling

Lecturer: David Whitmer *Scribes: Ilai Deutel, C.J. Argue*

1 Exponential Distributions

Definition 1.1. Given sample space Ω equipped with probability measure p, a **random variable** is a function $X : \Omega \to \mathbb{R}$

Definition 1.2. Given a random variable X, a **probability density function** (PDF) f_X for X satisfies:

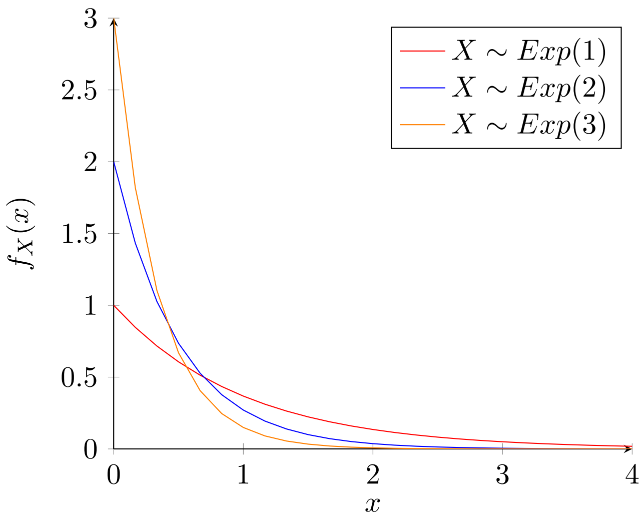
$$P(a \leq X \leq b) = \int_a^b f_X(x)\mathrm{d}x$$

Remark 1.3.

$$\int_{-\infty}^{+\infty} f_X(x)\mathrm{d}x = 1$$

Example 1.4. Exponential distribution with parameter β: $X \sim Exp(\beta)$

$$f_X(x) = \begin{cases} \beta \mathrm{e}^{-\beta x} & : \quad x \geq 0 \\ 0 & : \quad \text{otherwise} \end{cases}$$



Definition 1.5. The **cumulative distribution function** (CDF) for X is:

$$F_X = \Pr(X \leq x) = \int_{-\infty}^{x} f_X(x)\mathrm{d}x$$

Example 1.6. Exponential distribution with parameter β: $X \sim Exp(\beta)$

For $x \geq 0$,

$$F_X(x) = \int_{-\infty}^{x} \beta \mathrm{e}^{-\beta x}\mathrm{d}x = -\mathrm{e}^{-\beta t}\Big|_0^x = 1 - \mathrm{e}^{-\beta x}$$

Remark 1.7. In general, $f_X(x) = \dfrac{\mathrm{d}F_X(x)}{\mathrm{d}x}$

Definition 1.8. The **expected value** of X is:

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x f_X(x)\mathrm{d}x$$

Example 1.9. For an exponential distribution with parameter β: $X \sim Exp(\beta)$

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x\beta \mathrm{e}^{-\beta x}\mathrm{d}x = \frac{1}{\beta}$$

Proposition 1.10 (Memoryless Property of $Exp(\beta)$)**.** *If $X \sim Exp(\beta)$ then for any reals s, t, $\Pr[X \geq s + t \mid X \geq s] = \Pr[X \geq t]$*

Proof.

$$\begin{aligned}
\Pr[X \geq s + t \mid X \geq s] &= \frac{1 - \left(1 - \mathrm{e}^{-\beta(s+t)}\right)}{1 - (1 - \mathrm{e}^{-\beta s})} \\
&= \frac{\mathrm{e}^{-\beta(s+t)}}{\mathrm{e}^{-\beta s}} \\
&= \mathrm{e}^{-\beta t} \\
&= \Pr[X \geq t]
\end{aligned}$$

□

2 Order Statistics

Consider n i.i.d random variables $X_1, ..., X_n$.

Definition 2.1. $X_{(k)}$ is the k^{th} smallest of $X_1, ..., X_n$. That is, $X_{(i)}$ are a reordering of X_i s.t.

$$X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$$

Note that any single $X_{(i)}$ is a random variable, so it has a PDF f_X and CDF F_X.

Example 2.2. Let $X_i \sim Exp(\beta)$. In order to have $X_{(1)} \leq x$ we need to have $X_i \leq x$ (which has probability $F_X(x)$) for exactly one value of i and $X_i > x$ (probability $1 - F_X(x)$) for the remaining $n - 1$ values of i. There are n ways to choose which $X_i \leq x$, thus we have

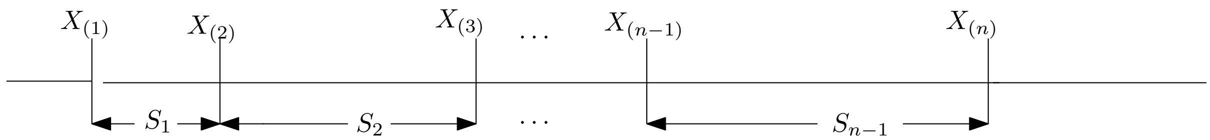
$$\begin{aligned}
f_{X_{(1)}}(x) &= f_X(x)(1 - F_X(x))^{n-1} n \\
&= \beta \mathrm{e}^{-\beta x(n-1)} n \\
&= n\beta \mathrm{e}^{-n\beta x}
\end{aligned}$$

Thus $X_{(1)} \sim Exp(n\beta)$, so $\mathbb{E}[X_{(1)}] = \dfrac{1}{n\beta}$.

We now establish the expectation of $X_{(n)}$.

Claim 2.3. $\mathbb{E}[X_{(n)}] = \Theta(\frac{\log n}{\beta})$.

Proof. Let $S_i = X_{(i+1)} - X_{(i)}$ for $i \geq 1$.



S_i is distributed as $\min(X_1, ..., X_{n-i}) - X_{(i)}$ conditioned on $\min(X_1, ..., X_{n-i}) \geq X_{(i)}$. By the memoryless property of exponential random variables, $[X_{(j)} \mid X_{(j)} \geq X_{(i)}] \sim X_{(j)}$. Thus S_i is the minimum of $n - i$ exponential random variables with parameter β, so by example 2.2, $S_i \sim Exp((n-i)\beta)$, and $\mathbb{E}[S_i] = \frac{1}{(n-i)\beta}$. Then:

$$\begin{aligned}
\mathbb{E}[X_{(n)}] &= \sum_{i=1}^{n-1} \mathbb{E}[S_i] + \mathbb{E}[X_{(1)}] \\
&= \sum_{i=1}^{n-1} \frac{1}{\beta(n-i)} + \frac{1}{n\beta} \\
&= \frac{1}{\beta}(1 + \frac{1}{2} + ... + \frac{1}{n}) \\
&\approx \frac{\log n}{\beta}
\end{aligned}$$

□

We now bound the variation of $X_{(n)}$.

Claim 2.4. $\Pr[X_{(n)} \geq \frac{2\log n}{\beta}] \leq \frac{1}{n}$

Proof.

$$\Pr\left[X_i \geq \frac{c\log n}{\beta}\right] = 1 - F_X\left(\frac{c\log n}{\beta}\right) = 1 - \left(1 - \mathrm{e}^{-\frac{c\log n}{\beta}\beta}\right) = n^{-c}$$

Thus we have

$$\Pr\left[X_{(n)} \geq \frac{c\log n}{\beta}\right] \leq n \Pr\left[X_i \geq \frac{c\log n}{\beta}\right] = n^{1-c}$$

Letting $c = 2$ gives $\Pr[X_{(n)} \geq \frac{2\log n}{\beta}] \leq \frac{1}{n}$. □

3 Sampling from distributions

3.1 Uniform distributions

Let U be the uniform distribution on $[0, 1]$.

- PDF: $f_U(x) = \begin{cases} 1 & : \quad x \in [0,1] \\ 0 & : \quad \text{otherwise} \end{cases}$
- CDF: $F_U(x) = \begin{cases} x & : \quad x \in [0,1] \\ 0 & : \quad x \leq 0 \\ 1 & : \quad x \geq 1 \end{cases}$

We will show how to use U to generate samples from other distributions.

Example 3.1. $Unif(0,2)$
$X \sim U \Rightarrow 2X \sim Unif(0,2)$

3.2 Inverse transform sampling

Say that we want to sample from random variable X and we have $Y \sim U$. We want to find some function g such that $g(Y)$ is distributed like X. To this end, note that if $g(Y)$ is an increasing function such that $F_{g(Y)} = F_X$ then

$$\begin{aligned} F_{g(Y)}(z) &= \Pr(g(Y) \leq z) \\ &= \Pr(Y \leq g^{-1}(z)) \\ &= g^{-1}(z) \end{aligned}$$

So $g^{-1}(z) = F_X(z) \Rightarrow g(z) = F_X^{-1}(z)$

Example 3.2. $X \sim Exp(\beta)$

$$F_X(x) = \begin{cases} 1 - \mathrm{e}^{-\beta x} & : \quad x \geq 0 \\ 0 & : \quad \text{otherwise} \end{cases}$$

To solve for F_X^{-1}

$$\begin{aligned} Y &= 1 - \mathrm{e}^{-\beta X} \\ \mathrm{e}^{-\beta X} &= 1 - Y \\ -\beta X &= \ln(1 - Y) \\ X &= -\frac{1}{\beta}\ln(1 - y) \end{aligned}$$

Thus

$$\begin{aligned} -\frac{\ln(1-Y)}{\beta} &\sim Exp(\beta) \\ -\frac{\ln(Y)}{\beta} &\sim Exp(\beta) \end{aligned}$$

The second line follows since $1 - Y$ and Y are both distributed as U.

3.3 Drawing from a normal distribution

A standard normal random variable $\mathcal{N}(0,1)$, mean 0, variance 1, is given by

- PDF: $f(x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}$
- CDF: $F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-t^2/2}\mathrm{d}t$

How do you sample from a normal distribution? The following theorem suggests that standard inverse transform sampling will not suffice.

Theorem 3.3. *$F(x)$ is not an elementary function*

3.3.1 Box-Muller sampling

Consider 2-dimensional normal (X, Y) ($X \sim \mathcal{N}(0,1)$, $Y \sim \mathcal{N}(0,1)$) with PDF:

$$\begin{aligned} f(x,y) &= \frac{1}{2\pi}\mathrm{e}^{-x^2/2}\mathrm{e}^{-y^2/2} \\ &= \frac{1}{2\pi}\mathrm{e}^{-\frac{x^2+y^2}{2}} \end{aligned}$$

In polar coordinates, $f(r,\theta) = \frac{1}{2\pi}\mathrm{e}^{-r^2/2}$ (circular/spherical symmetry: doesn't depend on θ).

Let R be the random variable $\sqrt{X^2+Y^2}$. The CDF $F_R(r)$ is given by

$$\begin{aligned} F_R(r) &= \Pr(R \le r) \qquad \text{(integral of } f(r,\theta) \text{ over disc of radius } r) \\ &= \int_0^{2\pi}\int_0^r \frac{1}{2\pi}\mathrm{e}^{-r'^2/2} r' \mathrm{d}r' \mathrm{d}\theta \\ &= \int_0^r \mathrm{e}^{-r'^2/2} r' \mathrm{d}r' \\ &= \int_0^{r^2} \frac{\mathrm{e}^{-u/2}}{2}\mathrm{d}u \qquad (\text{set } u = r'^2) \\ &= 1 - \mathrm{e}^{-r^2/2} \end{aligned}$$

We can now apply inverse transform sampling

$$\begin{aligned} Y &= 1 - \mathrm{e}^{-R^2/2} \\ R &= \sqrt{-2\ln(1-Y)} \end{aligned}$$

Thus $F_R^{-1}(y) = \sqrt{-2\ln(1-y)}$. To sample $X \sim \mathcal{N}(0,1)$, first draw from 2D normal (polar). Let $V, W \sim U$. Return (R, θ), where $R = \sqrt{-2\ln(V)}$ and $\theta = 2\pi W$. The rectangular coordinates, $(R\cos\theta, R\sin\theta)$, are independently distributed as $\mathcal{N}(0,1)$.

3.3.2 Polar form of Box-Muller

We can improve on this by eliminating the trigonometric functions, whose computation is costly. We will show an alternate way to sample normal variables via efficiently sampling a point from the uniform distribution on the unit disk. Consider generating a random variable as follows:

1. Choose (X, Y) uniformly at random on unit disk $\{X : \|X\| \le 1\}$
2. Let $S = X^2 + Y^2$
3. Return $\left(\sqrt{\frac{-2\ln S}{S}}X, \sqrt{\frac{-2\ln S}{S}}Y\right)$

We will show that this is equivalent to Box-Muller.

Claim 3.4. *If (X, Y) is chosen uniformly at random from the unit disk, then $X^2 + Y^2 \sim U$.*

Proof. It suffices to show that $\Pr[X^2 + Y^2 \le t] = t$

$$\begin{aligned}\Pr[X^2 + Y^2 \le t] &= \Pr[\sqrt{X^2 + Y^2} \le \sqrt{t}] \\ &= \frac{\text{Area of circle with radius } \sqrt{t}}{\text{Area of circle with radius 1}} \\ &= \frac{\pi t}{\pi} \\ &= t\end{aligned}$$

□

Also, symmetry implies that angle $\theta = \tan^{-1}(\frac{Y}{X})$ is uniform in $[0, 2\pi]$
So (X, Y) uniform on unit disk is distributed as $(\sqrt{V}\cos 2\pi W, \sqrt{V}\sin 2\pi W)$ for $V, W \sim U$.
$S = V$ and $\left(\sqrt{\frac{-2\ln S}{S}}X, \sqrt{\frac{-2\ln S}{S}}Y\right) = (R\cos\theta, R\sin\theta)$ above.

3.3.3 Sampling from the unit disk

To complete, we demonstrate how to generate uniform random point on unit disk without trigonometry. Intuitively, we throw darts at the square $[-1, 1] \times [-1, 1]$ until we hit the unit disk.

Let $V, W \sim U$, $X = 2U - 1$, $Y = 2U - 1$ so that X and Y are uniform on $[-1, 1]$. Let $S = X^2 + Y^2$.
Algorithm:

1. Draw V, W, calculate X, Y, S
2. If $S \le 1$, return (X, Y)
3. Else, try again

This is an example of **Rejection Sampling**. The probability of success in each iteration is $\frac{\text{Area of circle}}{\text{Area of square}} = \frac{\pi}{4}$, so in expectation we need to sample $\frac{4}{\pi} \approx \frac{4}{3}$ times.