QBS 120 - Lecture 2
Random variables
(Rice Chapter 2)

Rob Frost

Outline

- Discrete random variables
- Continuous random variables
- Functions of a random variable

Random variables (RVs)

Random variables are functions from Ω to the real numbers:

$$
\begin{aligned}
\Omega &= \{\omega_1, \dots, \omega_n\} \\
X &= \{x_1, \dots, x_n\}, x_i \in \mathbb{R} \\
X(\omega_i) &= x_i
\end{aligned}
$$

Probability measures on RVs are defined in terms of probability measure on Ω:

$$P_X(X = x_i) = P_\Omega(\omega_j \in \Omega : X(\omega_j) = x_i)$$

Example: coin toss

Fair coin is tossed three times

$$\Omega = \{hhh, hht, hth, thh, tth, tht, htt, ttt\}$$

RV X is number of heads

$$\begin{aligned} X(hhh) = 3, X(hth) = 2, \ldots \\ P_X(X = 3) = P_\Omega(hhh) = 1/8 \\ P_X(X = 1) = P_\Omega(\{tth, tht, htt\}) = 3/8 \end{aligned}$$

Discrete random variables

A discrete RV is an RV that can take on a finite or countably infinite number of values (one-to-one correspondence with integers).

Examples:

- Number of heads from prior example
  $x_i \in \{0, 1, 2, 3\}$
- Number of die rolls before rolling a 6
  $x_i \in \{1, 2, 3, ...\}$

Note: can also define discrete RVs in terms of CDF.

Probability mass function

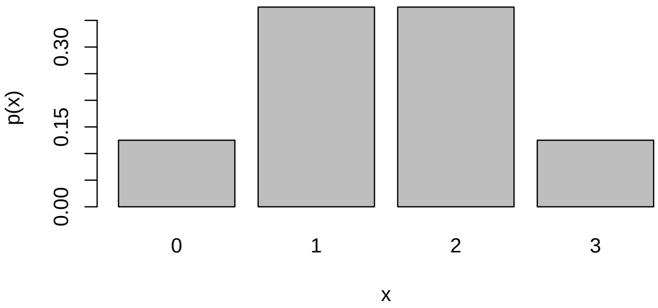
For discrete RVs, the probability function $P_X()$ is called the probability mass function (PMF) or frequency function $(P_X(X = x_i) = p(x_i))$

PMF properties:

- Assigns a valid probability to each potential x_i in the range of X, $0 \leq p(x_i) \leq 1$.
- Values for all potential x_i sum to 1, $\sum_i p(x_i) = 1$

PMF for coin toss example

```
> barplot(height=c(1/8, 3/8, 3/8, 1/8),
+         names.arg=0:3, xlab="x", ylab="p(x)")
```



Cumulative distribution function

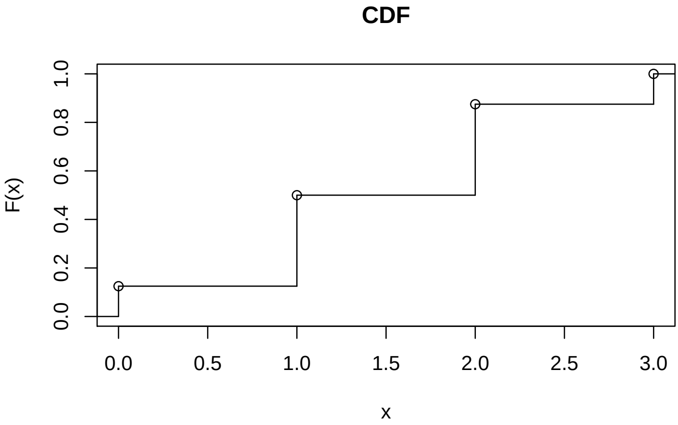
The cumulative distribution function (CDF) is a non-decreasing function that satisfies:

$$F(x) = P(X \leq x), -\infty < x < \infty$$
$$\lim_{x \to -\infty} F(x) = 0$$
$$\lim_{x \to \infty} F(x) = 1$$

For discrete RVs, the CDF is a step function with a jump at each x_i in the range of X.

CDF for coin toss example

```
> cdf.values = cumsum(c(1/8, 3/8, 3/8, 1/8))
> cdf = stepfun(0:3, c(0, cdf.values))
> plot(cdf, xlab="x", ylab="F(x)",
+         xlim=c(0,3),main="CDF")
```



Types of discrete RVs

- Bernoulli
- Binomial
- Geometric
- Negative binomial
- Hypergeometric
- Poisson

Bernoulli distribution

- Discrete RV with range $\{0, 1\}$
- PMF:

$$p(1) = p$$
$$p(0) = 1 - p$$
$$p(x) = 0, x \notin \{0, 1\}$$

- Alternate PMF representation:

$$p(x) = \begin{cases} p^x(1-p)^{1-x}, & x \in \{0, 1\} \\ 0, & x \notin \{0, 1\} \end{cases}$$

Bernoulli example: toss of one coin

- $\Omega = \{h, t\}$
- X is a Bernoulli RV with:

$$\begin{aligned} X(h) &= 1 \\ X(t) &= 0 \\ p &= 0.5 \end{aligned}$$

- Can also think of Bernoulli RVs as indicator functions (i.e., any indicator function is a Bernoulli RV):

$$I(\omega) = \begin{cases} 1, & \omega = h \\ 0, & \omega = t \end{cases}$$

Binomial distribution

- Discrete RV that represents the sum of n independent Bernoulli RVs all with the same probability p.
- *Example: A coin is tossed 10 times, the total number of heads is a binomial RV.*
- Possible values: $\{0, 1, ..., n\}$
- Binomial RVs have two parameters: n and p (Bernoulli RVs have just one, p)

Binomial PMF

What is probability that sum is k? (i.e., for binomial RV X, what is $P(X = k)$?)

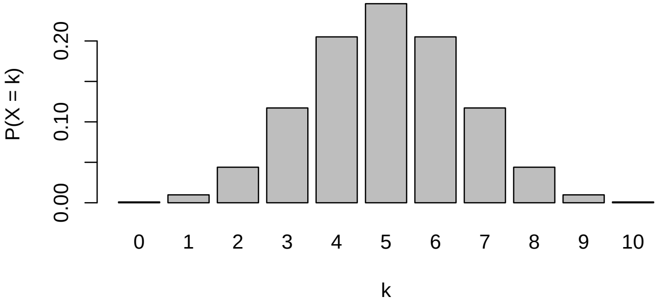
- By the multiplication principle, the probability of getting k 1s and $n - k$ 0s in n trials is $p^k(1-p)^{n-k}$.
- The number of unique ways to distribute k 1s across n trials is $\binom{n}{k}$, i.e., number the trials and then select k without replacement, the number of unordered such samples is the number of unique ways to assign k 1s to n trials.
- $P(X = k)$ is the probability of any one assignment times the number of assignments:

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

Binomial example

Coin is tossed 10 times (binomial RV with n=10, p=0.5)

```
> k.vals = 0:10
> probs = dbinom(k.vals, size=10, prob=0.5)
> barplot(height=probs, names.arg=k.vals,
+     xlab="k", ylab="P(X = k)")
```



R support for random variables

- d*(): density or pmf
- p*(): CDF
- q*(): quantile (inverse CDF)
- r*(): generates random values

```
> dbinom(4, size=10, prob=0.5)
[1] 0.2050781
> (p  = pbinom(4, size=10, prob=0.5))
[1] 0.3769531
> qbinom(p, size=10, prob=0.5)
[1] 4
> rbinom(5, size=10, prob=0.5)
[1] 6 3 6 4 4
```

Geometric distribution

- Discrete RV that is also based on a sequence of independent Bernoulli RVs.
- For geometric RVs, care about the number of trials before encountering the first 1.
- Range: $\{1, \ldots, \infty\}$, just one parameter p.
- PMF (from multiplication principle)

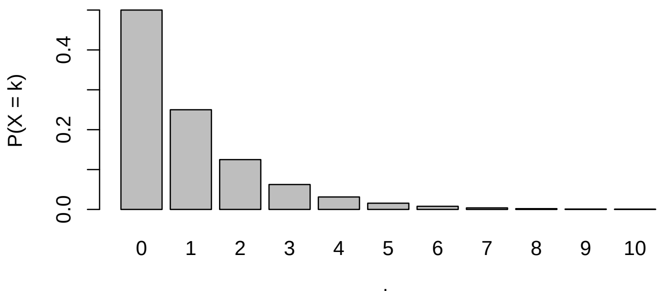
$$p(k) = P(X = k) = (1 - p)^{k-1} p$$

Note: can show that $\sum_{i=1}^{\infty} (1 - p)^{k-1} p = 1$ *using properties of geometric series.*

Geometric example

How many coin tosses to get the first heads? ($p = 0.5$)
Note that the k values for dgeom() are number of failures before first success so 0 is part of range.

```
> k.vals = 0:10
> probs = dgeom(k.vals, prob=0.5)
> barplot(height=probs, names.arg=k.vals,
+     xlab="k", ylab="P(X = k)")
```



Negative binomial distribution

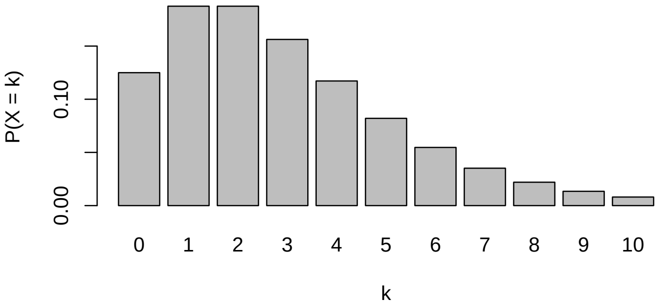
- Generalization of the geometric distribution.
- Care about the number of Bernoulli trials until there are r 1s.
- Range: $\{1, ..., \infty\}$, two parameters: p and r
- PMF
  - Probability of any particular sequence of r 1s in k trails $=$ $p^r(1-p)^{k-r}$
  - Since last trial must be a 1, have $\binom{k-1}{r-1}$ ways to assign remaining 1s.

$$p(k) = P(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$$

Negative binomial example

How many coin tosses to get 3 heads? ($p = 0.5, r = 3$)

```
> k.vals = 0:10
> probs = dnbinom(k.vals, size=3, prob=0.5)
> barplot(height=probs, names.arg=k.vals,
+    xlab="k", ylab="P(X = k)")
```



Hypergeometric distribution

- If sampling without replacement is done on a population of n items with r of one type and $n-r$ of the other type (e.g., urn of n balls with r black and $n-r$ white), how many items of the first type are included among m draws without replacement?
- Range $\{0, \ldots, r\}$, parameters r, n, m.
- PMF

$$P(X=k) = \frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}}$$

Important for contingency table analysis.

Poisson distribution

- Poisson distribution approximates the binomial distribution for large n and small p.
- Effective model for the count of rare events in a given time interval (or spatial region).
- Motivation: computational complexity. Although not a major issue today, the Poisson PMF is much easier to compute than the binomial for large n. Also has some nice features for estimation.

Poisson PMF

Can derive the Poisson PMF as the limit of the bionomial PMF as $n \to \infty, p \to 0, np = \lambda$.

$$P(X = k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \qquad \text{binomial pmf}$$

$$P(X = k) = \frac{n!}{k!(n-k)!} \frac{\lambda}{n}^k (1 - \frac{\lambda}{n})^{n-k} \qquad np = \lambda$$

$$P(X = k) = \frac{\lambda^k}{k!} \frac{n!}{(n-k)!n^k} (1 - \frac{\lambda}{n})^n (1 - \frac{\lambda}{n})^{-k}$$

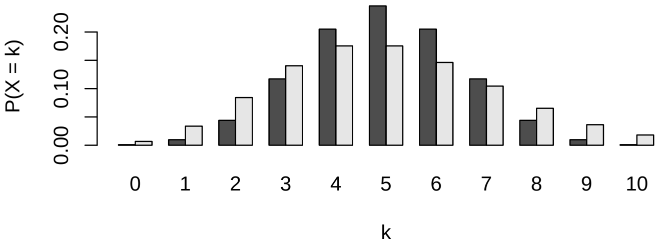
$$\text{as } n \to \infty : \frac{n!}{(n-k)!n^k} \to 1, (1 - \frac{\lambda}{n})^n \to e^{-\lambda}, (1 - \frac{\lambda}{n})^{-k} \to 1$$

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Poisson example

Coin is tossed 10 times, how many heads? (binomial RV with n=10, p=0.5) How is Poisson approximation?

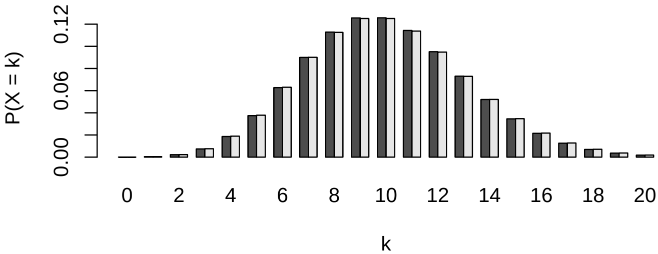
```
> k.vals = 0:10
> probs = dbinom(k.vals, size=10, prob=0.5)
> probs.pois = dpois(k.vals, lambda = (.5*10))
> barplot(height=rbind(probs,probs.pois),
+     beside=T, names.arg=k.vals,
+     xlab="k", ylab="P(X = k)")
```



Poisson example, 2

1k people are exposed to infectious agent with chance of infection of 1%. How many are infected? (binomial RV with n=1000, p=0.01). How is Poisson approximation?

```
> k.vals = 0:20
> probs = dbinom(k.vals, size=1000, prob=0.01)
> probs.pois = dpois(k.vals, lambda = 10)
> barplot(height=rbind(probs,probs.pois),
+    beside=T, names.arg=k.vals,
+    xlab="k", ylab="P(X = k)")
```



Outline

- Discrete random variables
- **Continuous random variables**
- Functions of a random variable

Continuous random variables

A continuous RV is an RV that can take on a continuum of values (e.g., any real number between 0 and 1).

The density function, $f(x)$, takes the place of the PMF. Properties of $f(x)$:

- $f(x) \geq 0$ for all x in range of X.
- $f(x)$ is piecewise continuous
- $\int_{-\infty}^{\infty} f(x)dx = 1$

Probabilities are found by integration:

$$P(a < X < b) = \int_a^b f(x)dx$$

What is $P(X = x)$ for a continuous RV X?

What is $P(X = x)$ for a continuous RV X?

$$P(X = x) = \int_x^x f(x)dx = 0$$

$$P(a < X < b) = P(a \leq X \leq b)$$

CDF for continuous RVs

CDFs for continuous and discrete RVs have the same definition:

$$F(x) = P(X \leq x)$$

For continuous RVs, F(x) can be evaluated via integration of density function:

$$F(x) = \int_{-\infty}^{x} f(x)dx$$

Quantiles

Inverse of $F(x)$ gives quantiles for the RV X:

$$F(x_p) = p \rightarrow P(X \leq x_p) = p$$
$$x_p = F^{-1}(p)$$

x_p is referred to as the pth quantile of X; $x_{0.5}$ is the median.

```
> (p = pbinom(3, size=10, prob=0.5))
[1] 0.171875
> (x_p = qbinom(p, size=10, prob=0.5))
[1] 3
```

Types of continuous RVs

- Uniform
- Exponential
- Gamma
- Normal
- Beta

Uniform distribution

- Represents a random real number on the interval $[a, b]$, i.e., all numbers in the interval are equally likely.
- Density function:

$$f(x) = \begin{cases} 1/(b-a), & a \leq x \leq b \\ 0, & x < a \text{ or } x > b \end{cases}$$

- CDF:

$$F(x) = \begin{cases} 0, & x < a \\ (x-a)/(b-a), & a \leq x \leq b \\ 1, & x > b \end{cases}$$

Standard uniform distribution

- Uniform distribution on the interval $[0,1]$ (what is approximated by most pseudo random number generators; model for p-values under H_0)
- Density function:

$$f(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & x < 0 \text{ or } x > 1 \end{cases}$$

- CDF:

$$F(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \leq x \leq 1 \\ 1, & x > 1 \end{cases}$$

Exponential distribution

- Continuous distribution used to model lifetimes, e.g., survival time of an organism, lifetime of a electromechanical device, etc.
- Like Poisson, has a single parameter λ that controls the rate at which the density declines as x increases.
- Density function:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

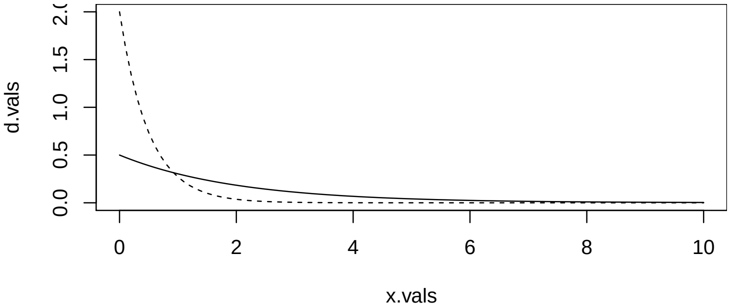
- CDF (compute via integration of $f(x)$):

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Exponential distribution, continued

Density functions for $\lambda = 0.5$ (solid) and $\lambda = 2$ (dashed):

```
> par(mar=c(4,4,0,0))
> x.vals = seq(from=0,to=10,by=0.1)
> d.vals = dexp(x.vals, rate=2)
> plot(x.vals, d.vals, type="l", x.lab="x", y.lab="f(x)",
+         lty="dashed")
> d.vals = dexp(x.vals, rate=0.5)
> lines(x.vals, d.vals, type="l")
```



Exponential distribution, continued

The exponential distribution has the "memoryless" property:

- Example: modeling lifetime of mouse as exponential RV T.
- If the mouse is alive at time s, what is the probability it will live an additional time t?

$$
\begin{aligned}
P(T > t + s | T > s) &= \frac{P(T > t + s \cap T > s)}{P(T > s)} \\
&= \frac{P(T > t + s)}{P(T > s)} \\
&= \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} \\
&= e^{-\lambda t}
\end{aligned}
$$

Probability does not depend on s! A good model for animal survival?

Gamma distribution

- Generalization of the exponential (and χ^2).
- Parameters: α and λ.
- Density function:

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

- Reduces to exponential when $\alpha = 1$.
- Can represent many types of non-negative RVs.

Normal distribution (Gaussian)

- Key statistical distribution due to Central Limit Theorem: sum of independent RV is approximately normal.
- Parameters: μ (mean) and σ (standard deviation)
- Density function:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, -\infty < x < \infty$$

Question: can we identify the formula for the CDF via integration of the density?

Normal distribution, continued

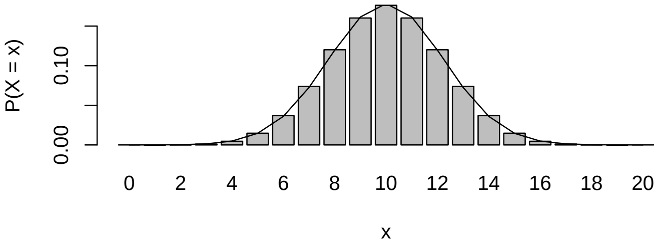
Answer: No!

The normal density does not have a closed-form integral. It is referred to as the error function (erf, $\int e^{-x^2}$), a *special function* that must be evaluated numerically.

Normal example

Number of heads in 20 coin tosses. We know this has a binomial distribution with $p = 0.5$ and $n = 20$. Can it be approximated by a normal distribution?

```
> x.vals = 0:20
> binom.probs = dbinom(x.vals, p=0.5, size=20)
> norm.probs = dnorm(x.vals, mean=10, sd=sqrt(5))
> df.bar = barplot(height=binom.probs, names.arg=x.vals,
+     xlab="x", ylab="P(X = x)")
> lines(df.bar, norm.probs, type="l")
```



Beta distribution

- Used to model continuous RVs restricted to $[0, 1]$.
- Parameters a and b
- Density function:

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}, 0 \leq x \leq 1$$

- Important in Bayesian analysis.

Random variable recap

RVs are functions that map from sample space to real numbers:

$$\Omega \rightarrow \mathbf{X}(\omega_\mathbf{i}) \rightarrow x_i \in \mathbb{R}$$
$$\{hhh, hht, ..., ttt\} \rightarrow \mathbf{X}(\mathbf{hhh}) \rightarrow 3$$

PMF/density functions map from range of RV to probabilities:

$$x_i \rightarrow \mathbf{P}(\mathbf{X} = \mathbf{x_i}) \rightarrow p_i, 0 \leq p_i \leq 1$$

Outline

- Discrete random variables
- Continuous random variables
- **Functions of a random variable**

Functions of RVs

If RV X has density function $f(x)$ and CDF $F(x)$, what are the density and CDF of $Y = g(X)$?

Linear functions of RVs

$Y = aX + b$. What is CDF of Y?

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(aX + b \leq y) \\ &= P(X \leq \frac{y-b}{a}) \\ &= F_X(\frac{y-b}{a}) \end{aligned}$$

What is density of Y?

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_X(\frac{y-b}{a}) \\ &= \frac{1}{a} f_X(\frac{y-b}{a}) \end{aligned}$$

Linear functions of normal RVs

$X \sim \mathcal{N}(\mu, \sigma^2),\ Y = aX + b$

$$
\begin{aligned}
f(y) &= \frac{1}{a} f_X\left(\frac{y-b}{a}\right) \\
&= \frac{1}{a\sigma\sqrt{2\pi}} e^{-((y-b)/a-\mu)^2/2\sigma^2} \\
&= \frac{1}{a\sigma\sqrt{2\pi}} exp\Big[-\frac{1}{2}\Big(\frac{y-b-a\mu}{a\sigma}\Big)^2\Big]
\end{aligned}
$$

$Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$

Standard normal equivalence

A key application of that result is representing properties of arbitrary normal RVs ($X \sim \mathcal{N}(\mu, \sigma)$) in terms of the standard normal distribution ($Z \sim \mathcal{N}(0, 1)$).

- Create a linear transformation of X: $Z = (X - \mu)/\sigma$ ($a = 1/\sigma$, $b = -\mu/\sigma$)
- Apply the property regarding linear transformations of normal variables to find that:

$$\begin{aligned} Z &\sim \mathcal{N}(a\mu + b, a^2\sigma^2) \\ &\sim \mathcal{N}(\mu/\sigma - \mu/\sigma, (1/\sigma^2)\sigma^2) \\ &\sim \mathcal{N}(0, 1) \end{aligned}$$

Standard normal equivalence, continued

To find CDF of X:

$$
\begin{aligned}
F_X(x) &= P(X \leq x) \\
&= P(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}) \\
&= P(Z \leq \frac{x-\mu}{\sigma}) \\
&= \Phi(\frac{x-\mu}{\sigma})
\end{aligned}
$$

To find probability that $x_0 < X < x_1$:

$$
\begin{aligned}
P(x_0 < X < x_1) &= F_X(x_1) - F_X(x_0) \\
&= \Phi(\frac{x_1-\mu}{\sigma}) - \Phi(\frac{x_0-\mu}{\sigma})
\end{aligned}
$$

Chi-square (χ^2) example

What is the density of $X = Z^2$ where $Z \sim \mathcal{N}(0,1)$?

In this case, $g()$ is not linear so can't apply prior results but we'll use a similar procedure (i.e., find CDF and differentiate to get density).

$$\begin{aligned}
F_X(x) &= P(X \leq x) \\
&= P(-\sqrt{x} \leq Z \leq \sqrt{x}) \\
&= \Phi(\sqrt{x}) - \Phi(-\sqrt{x})
\end{aligned}$$

To get density, differentiate the CDF:

$$\begin{aligned}
f_X(x) &= \frac{\delta}{\delta x}(\Phi(\sqrt{x}) - \Phi(-\sqrt{x})) \\
&= 1/2x^{-1/2}\phi(\sqrt{x}) + 1/2x^{-1/2}\phi(-\sqrt{x}) && \text{apply chain rule} \\
&= x^{-1/2}\phi(\sqrt{x}) && \phi \text{ is symmetric} \\
&= \frac{x^{-1/2}}{\sqrt{2\pi}} e^{-x/2}
\end{aligned}$$

A gamma density with $\alpha = \lambda = 1$ or the chi-square density.

CDFs and uniform RVs

- The RV A is set equal the CDF of another RV X: $A = F(X)$
- What is the distribution of A?

$$\begin{aligned} P(A \leq a) &= P(F(X) \leq a) \\ &= P(X \leq F^{-1}(a)) \\ &= F(F^{-1}(a)) \\ &= a \end{aligned}$$

- A therefore has a uniform distribution on $[0, 1]$

CDFs and uniform RVs, continued

- The RV A is set equal the inverse of CDF $F()$ applied to a standard uniform RV U: $A = F^{-1}(U)$
- What is the CDF of A?

$$
\begin{aligned}
P(A \leq a) &= P(F^{-1}(U) \leq a) \\
&= P(U \leq F(a)) \\
&= F(a)
\end{aligned}
$$

- A therefore has CDF $F()$.

Generation of pseudo-random numbers

Can use the result on the previous slide to generate pseudo-random numbers following a desired distribution from standard uniform pseudo-random numbers (what is typically produced by computer pseudo-random number generators).

- Generate a pseudo-random number in [0,1].
- Apply inverse of desired CDF to the uniform random number.