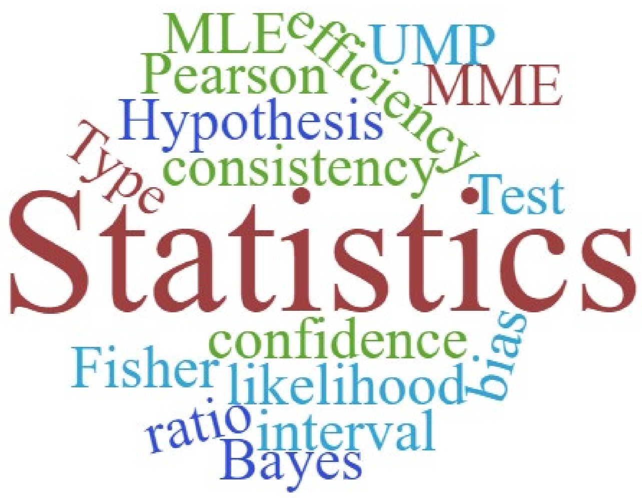
Math 181A: Mathematical Statistics



David Quarfoot
Associate Teaching Professor
Email: dquarfoot@ucsd.edu
Office: APM 5220

Learning Goals:

- Review the four most common discrete distributions and four most common continuous distributions (for each, memorize the idea it measures, support, pmf/pdf, expected value, and variance)
- Practice using these distributions in real-life problems
- Introduce the d-p-q-r quartet of functions in R

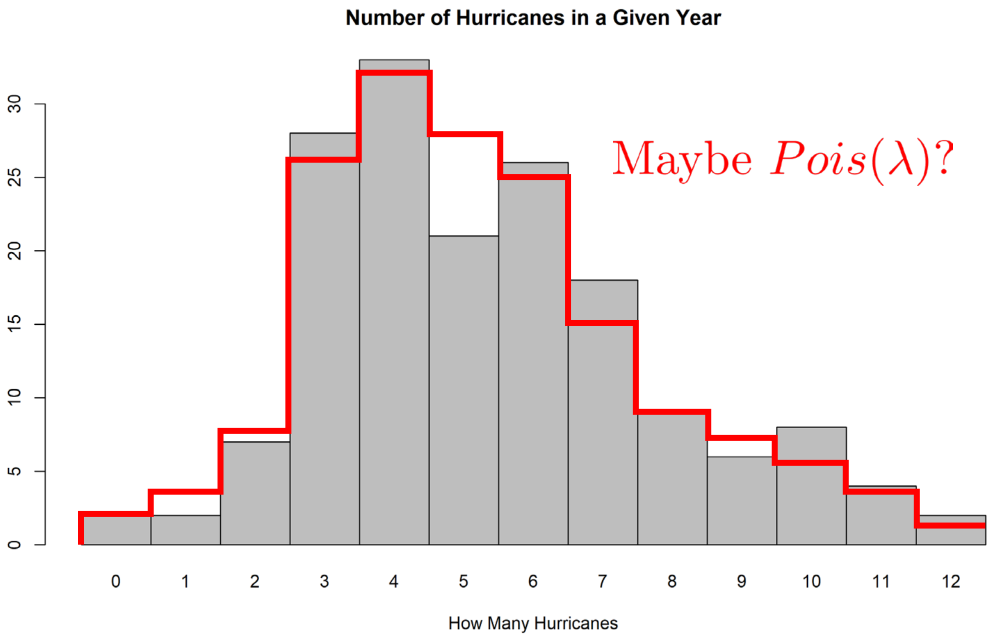
Welcome to Mathematical Statistics!

Calculus : Real Analysis :: Statistics : Mathematical Statistics

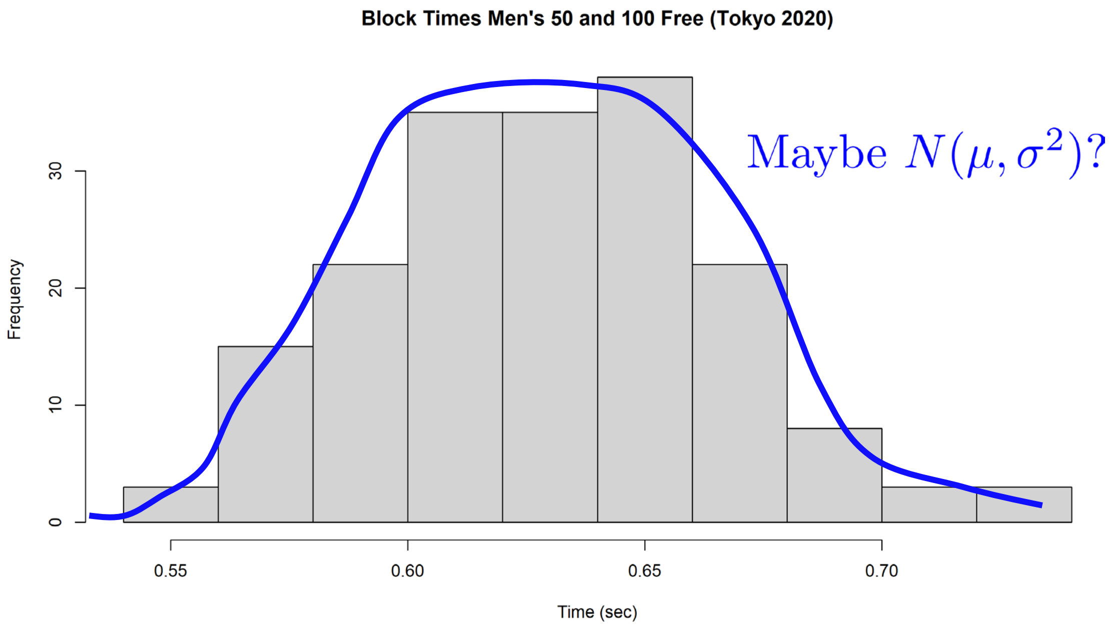
Our goal is to rigorously derive (when possible) many of the main results from a traditional statistics class and to focus on the mathematics that underpin this beautiful subject.

Main topics: Parameter estimation, properties of estimators, hypothesis testing, confidence intervals, errors and decision rules, likelihood ratios, simulation using computers

Modeling 101



discrete random variable (DRV)



continuous random variable (CRV)

In each case, we must use our *data and knowledge of the situation* to choose an appropriate statistical model: $Pois(\lambda)$, $N(\mu, \sigma^2)$. The data can help us determine the best parameter values (e.g., $\lambda = 5.44$ and $\mu = 0.641, \sigma^2 = 0.002$).

Summary of Our Models: Discrete Models

Geometric: $X \sim Geom(p)$

$$X \in \{1, 2, 3, \ldots\} \quad (\text{“support”})$$

X is the number of trials needed to get the first success (including the success). Each trial has success probability p.

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

Sometimes we'll write $P(X = k)$, or simply $P(k)$ if the random variable (RV) is clear.

$$E(X) = \frac{1}{p}, \; Var(X) = \frac{1-p}{p^2} = \frac{q}{p^2}$$

Binomial: $X \sim Binom(n, p)$

$$X \in \{0, 1, 2, \ldots, n\}$$

X is the number of successful trials of out the n total trials. Each trial has success probability p.

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

$$E(X) = np, \; Var(X) = np(1-p)$$

It will help later if you have these memorized without using $q = 1 - p$, since fewer variables appear visually.

This info is also on pages 274 – 276 of Larsen and Marx, 6th edition.

Poisson: $X \sim Poisson(\lambda)$

$$X \in \{0, 1, 2, \ldots\}$$

X is the number of times an event occurs in a given time when its average rate of occurrence in that time is λ.

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

$$E(X) = Var(X) = \lambda$$

Note that the units on λ (rate) are:

$$\frac{\text{number of events}}{\text{time span for measurement}}$$

Negative Binomial: $X \sim NegBinom(r, p)$

$$X \in \{r, r+1, r+2, \ldots\}$$

X is the number of trials needed to get the r^{th} success (so r total successes).
Each trial has success probability p.

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

$$E(X) = \frac{r}{p}, \; Var(X) = \frac{r(1-p)}{p^2}$$

When $r = 1$, this collapses to the Geometric distribution.
Indeed, $NegBinom(r, p)$ is just a sum of r $Geom(p)$ RVs.
This is how I remember the $E(X)$ and $Var(X)$ formulas.

Exams are closed notes/books and no cheat sheets. Memorize these today.

Your Turn: Bad Service

You will count the number of tennis serves needed for you to get your first ball in. Your friend says it's bad luck if it takes an odd number of attempts. If your serves are independent and each has probability 0.01 of being in, what's the probability you have bad luck?

Let X be the number of attempts to get the first success (ball served in).
We know $X \sim Geom(0.01)$. Always explain your RV in words and give its distribution.

We want $p_{\text{bad luck}} = P_X(X \text{ is odd}) = P(X = 1) + P(X = 3) + P(X = 5) + \ldots$

Recall that: $P_X(X = k) = (1 - p)^{k-1}p$

$$= 0.01 + (0.99)^2(0.01) + (0.99)^4(0.01) + \ldots$$

sum of infinite geom. series $= \dfrac{\text{first term}}{1 - \text{common ratio}}$

$$= \frac{0.01}{1 - (0.99)^2} = \boxed{\frac{100}{199} \approx 0.503}.$$

In this class, give exact answers or round to three decimal places.

Summary of Our Models: Continuous Models

Uniform: $X \sim Unif(a,b)$

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & \boxed{a \le x \le b} \text{ ("support")} \\ 0, & \text{else} \end{cases}$$

X is an output in a finite range for which all outcomes in the range are equally likely.

$$E(X) = \frac{a+b}{2},\ Var(X) = \frac{(b-a)^2}{12}$$

Recall: All density functions satisfy:

1) $f(x) \ge 0$ for all values of x

2) $\int_{-\infty}^{\infty} f(x)\,dx = 1$

Exponential: $X \sim Exp(\lambda)$

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \ge 0 \\ 0, & \text{else} \end{cases}$$

X represents how long we must wait for an event to occur when we know how often it occurs on average (λ).

$$E(X) = \frac{1}{\lambda},\ Var(X) = \frac{1}{\lambda^2}$$

Note that the units on λ (rate) are:

$$\frac{\text{number of events}}{\text{time span for measurement}}$$

This is the same setup for λ as Poisson.

Normal: $X \sim N(\mu, \sigma^2)$

In this class, the number after the comma will be the variance.

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

X represents a real-world phenomenon that follows a bell-curve distribution.

$$E(X) = \mu, \ Var(X) = \sigma^2$$

Gamma: $X \sim Gamma(r, \lambda)$

$$f(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}, x > 0, r > 0, \lambda > 0$$

X represents the time required to see r instances of an event that occurs with average rate λ.

If $r \in \mathbb{N}$, then $Gamma(r, \lambda) = \sum_{1}^{r} Exp(\lambda)$

The λ will remind you of the exponential dist.; the r will remind you of repeated ideas.

Memory aid – Geom : NegBinom :: Exp : Gamma

$$E(X) = \frac{r}{\lambda}, \ Var(X) = \frac{r}{\lambda^2}$$

These two models have the hardest-to-memorize density functions. Spend extra time learning them.

Understanding The Gamma Distribution

$$f(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}, x > 0$$

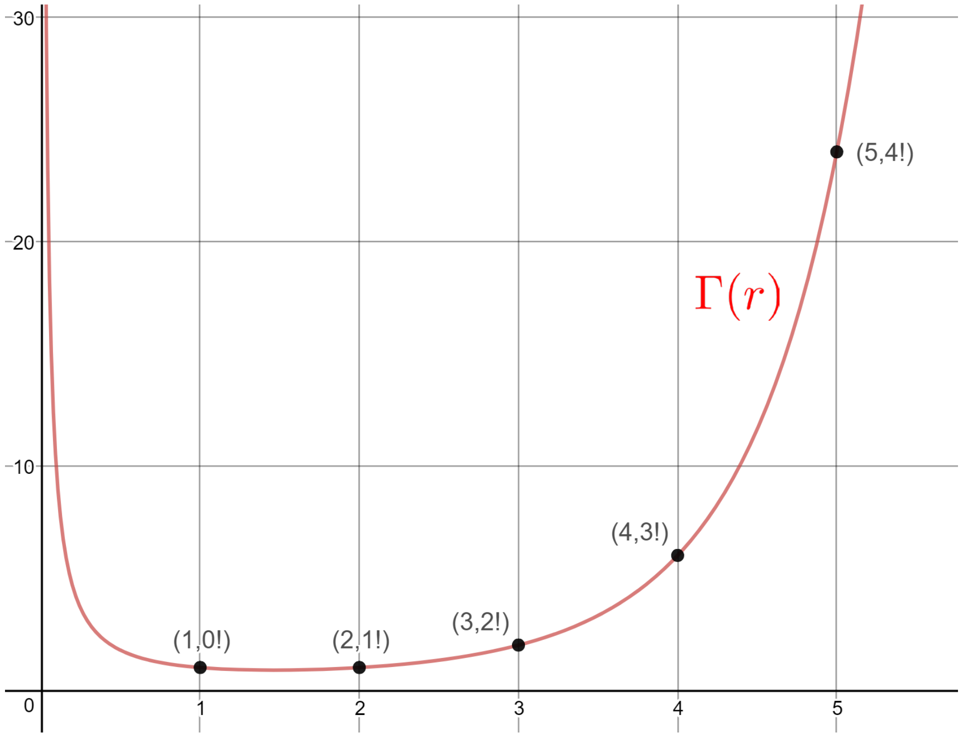
The gamma *function*, $\Gamma(r)$, is a generalization of the factorial function for non-natural numbers:

$$\Gamma(r) = \int_0^\infty y^{r-1} e^{-y}\, dy$$

Important facts to remember:

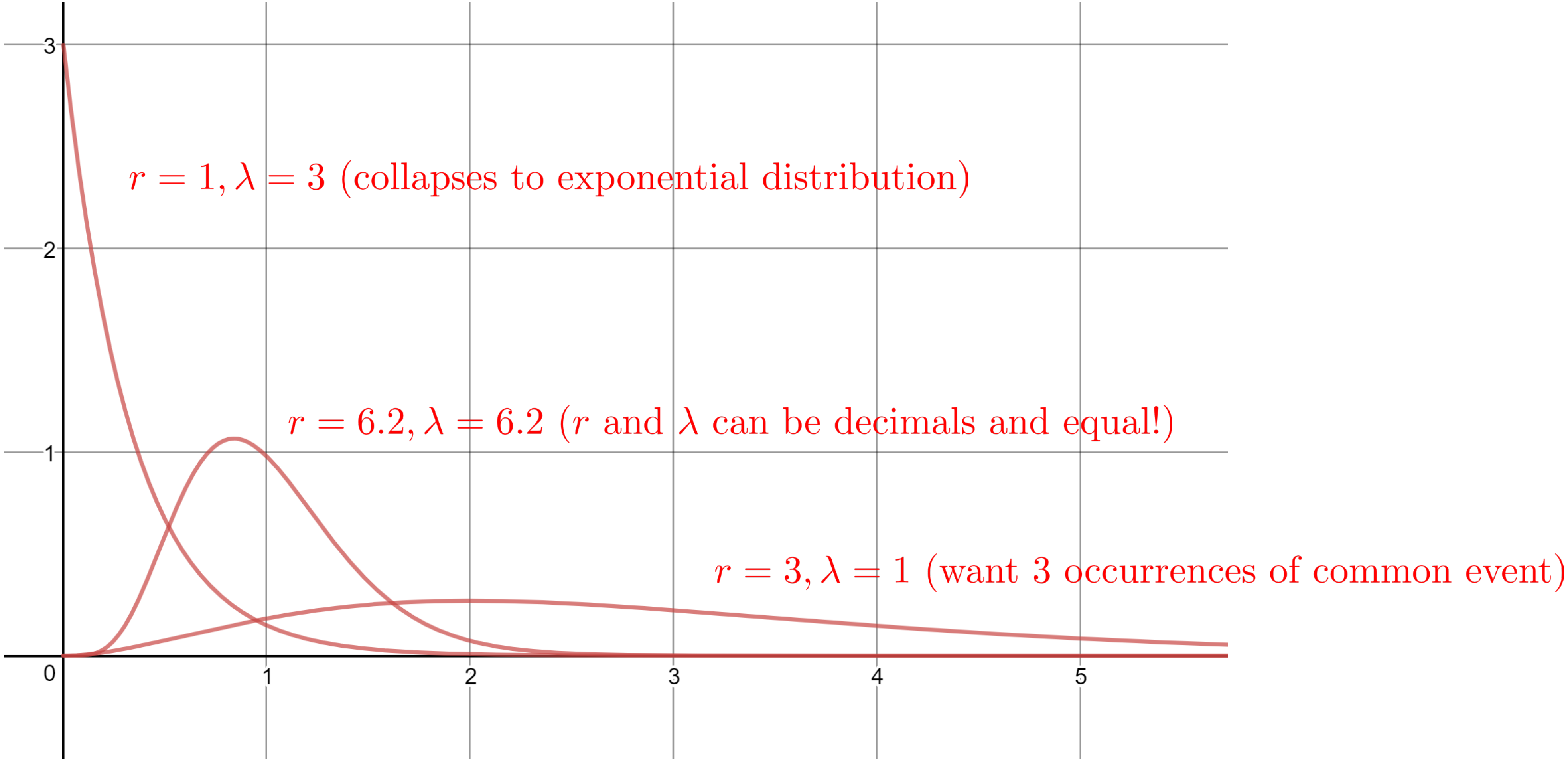
- $\Gamma(r)$ is a number
- $\Gamma(x+1) = x \cdot \Gamma(x)$ for any $x > 0$
- $\Gamma(r) = (r-1)!$ if $r \in \mathbb{N}$

Note: 0! and 1! are defined to be 1.
Thus, $\Gamma(2) = 1! = 1$ and $\Gamma(1) = 0! = 1$.



Visualizing the Gamma Distribution

Explore in Desmos [here]



Additional Review Ideas

You should explore the problems that follow on your own to deepen your review of 180A. These slides refresh rules for expectation, rules for variance, and introduce some basic commands in R.

Your Turn: Red Planet

You are designing the energy system for the next Mars rover, which is required to function for 2 years. Your team has designed a new solar panel that breaks, on average, once every 8 months. If you equip the rover with 3 such panels, and a new one turns on after an old one breaks, what is the probability you will be able to power the mission?

Let $X \sim Exp\left(\lambda = \dfrac{1 \text{ failure}}{2/3 \text{ years}}\right)$ model the time-til-failure for a single panel.

Let Y be the time-til-failure for the three-panel system.

We know $Y = X_1 + X_2 + X_3 \sim Gamma(r = 3, \lambda = 3/2)$.

The problem wants: $P(Y > 2) = \int_2^\infty \frac{(3/2)^3}{2!} x^2 e^{-3x/2}\, dx$

We could do this integral by hand (integration by parts twice) or have R help us calculate it!

R has built-in functions for all the major distributions, and for each, a quartet of functions (the d-p-q-r quartet) that allow you to interact with the distribution:

- dgamma: Find the height of the gamma density function at some place
- pgamma: Find an area (probability) under gamma up to some place
- qgamma: Find quantiles related to the gamma function
- rgamma: Generate random values from the gamma function

```
> pgamma(2,shape = 3,rate = 3/2, lower = F)
[1] 0.4231901
```

"shape" plays the role of r
"rate" plans the role of λ
"lower = F" means we shade toward the upper tail

Your Turn: Joining Forces

You plan to throw 12 coins each with $p_{\text{heads}} = \dfrac{2}{3}$ and count the number of heads, X. You also plan to flip a single one of these coins until you get a total of 8 tails and record the number of flips, Y. What is the expected value of $2X - 4Y + 3$?

Note that $X \sim Binom(n = 12, p = 2/3)$ and $Y \sim NegBinom(r = 8, p = 1/3)$.

Thus, $E(X) = np = 12 \cdot \dfrac{2}{3} = 8$ and $E(Y) = \dfrac{r}{p} = \dfrac{8}{1/3} = 24$.

Recall: $E(X+Y) = E(X)+E(Y)$, $E(X+c) = E(X)+c$, and $E(aX) = aE(X)$.
These formulas are true if X and Y are dependent or independent RVs.

By the linearity of the expected value (see red above), we get:
$E(2X - 4Y + 3) = 2E(X) - 4E(Y) + 3 = 2 \cdot 8 - 4 \cdot 24 + 3 = \boxed{-77}$.

Welcome to the Monte Carlo Casino!

A "Monte Carlo simulation" is simply using the computer to answer a statistical question by generating a lot of data. This is useful if a problem is too difficult to solve theoretically.

Example (by-hand and via simulation): Suppose X is from $Unif(-1,1)$ and W is from $N(0,4)$. How much variability is there in the expression $Y = 3X - W + 2$?

Recall: $Var(aX) = a^2 Var(X)$, $Var(X \pm Y) = Var(X) + Var(Y)$ if X and Y are independent RVs, and $Var(X + c) = Var(X)$.

$$Var(Y) = Var(3X - W + 2) = Var(3X - W) = Var(3X) + Var(W) = 9Var(X) + Var(W)$$

Using the variances for these distributions, we get $9 \cdot \dfrac{(1-(-1))^2}{12} + 4 = \boxed{7}$.

R Casino

Let's try to simulate this variability.

Make 1000 random numbers from $Unif(-1, 1)$ and look at them for fun (type the variable name!)

```
> X = runif(n = 1000, min = -1, max = 1)
> X
   [1]  0.713052417 -0.327821328  0.040786834 -0.637410878  0.580789912  0.127814044
   [7]  0.820366843  0.846911240 -0.418139182 -0.124171207  0.473739811  0.886364816
  [13] -0.254688424 -0.960759468  0.602528248 -0.983886629  0.013596401 -0.792400826
  [19] -0.255214440  0.595800061 -0.624271540  0.414714573 -0.259655388 -0.295617895
  [25] -0.008395093 -0.367478200  0.848243620 -0.185383519 -0.549397314 -0.684977041
```

R can do "vectorized math"

```
> 3*X
   [1]  2.139157251 -0.983463984  0.122360503 -1.912232635  1.742369737  0.383442131
   [7]  2.461100529  2.540733721 -1.254417545 -0.372513620  1.421219432  2.659094449
  [13] -0.764065271 -2.882278404  1.807584744 -2.951659887  0.040789202 -2.377202477
```

Even more "vectorized math"

```
> Y = 3*X - rnorm(n = 1000, mean = 0, sd = 2) + 2
> Y
   [1]  5.70434978 -1.17712490 -1.26667805 -0.97018060  2.44757387 -1.09676367
   [7]  4.62271971  4.99986462  2.87987775 -2.20001883  3.50464831  7.89290125
  [13]  1.31311287 -4.38925236  2.93719237 -1.81807020  5.17465035 -1.66094282
  [19] -2.53406056  5.94139952  2.36613919 -0.17376622  2.97417217  0.29522404
```

```
> var(Y)
[1] 6.950089
```

Don't expect your result to perfectly match the theoretical value (if you know one). Real-life data are noisy. How could we get the result to be closer to 7?