From last time

note: $X_{1:n} = \{X_1, X_2, \ldots, X_n\}$

Data: $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Exp}(\lambda)$

joint density of the data.

likelihood function:

$$L(\lambda \mid X_{1:n}) = f(X_1, X_2, \ldots, X_n \mid \lambda)$$

iid

$$= \prod_{i=1}^{n} f(X_i \mid \lambda)$$

$$= \prod_{i=1}^{n} \lambda e^{-\lambda X_i}$$

$$= \lambda^n e^{-\lambda \sum_{i=1}^{n} X_i}$$

$$\ell(\lambda \mid X_{1:n}) = \ln L(\lambda \mid X_{1:n}) = n \ln \lambda - \lambda \sum_{i=1}^{n} X_i$$

MLE: $$\hat{\lambda}_n^{(MLE)} = \underset{\lambda > 0}{\text{argmax}}\, L(\lambda \mid X_{1:n}) = \underset{\lambda > 0}{\text{argmax}}\, \ell(\lambda \mid X_{1:n}) = \frac{1}{\bar{X}_n}.$$

properties: with math . . .

$$\text{bias}\left(\hat{\lambda}_n^{(MLE)}\right) = E\left(\hat{\lambda}_n^{(MLE)}\right) - \lambda$$

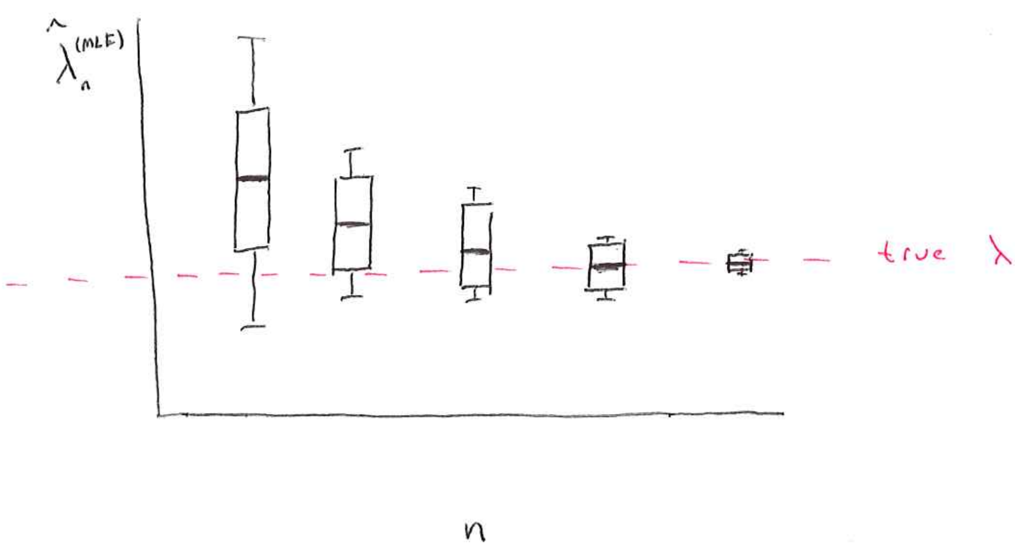
$$= \frac{n\lambda}{n-1} - \lambda$$

$$= \frac{\lambda}{n-1} > 0.$$

biased upward!

with simulation . . . $\hat{\lambda}_n^{(MLE)} \longrightarrow \lambda$ as $n \longrightarrow \infty$.

Consistent!



$100 \times (1-\alpha)\%$ confidence intervals for λ

Use the data to compute bounds

$$L_n = \text{lower}(X_{1:n})$$

$$U_n = \text{upper}(X_{1:n})$$

satisfying $P(L_n < \lambda < U_n) = 1 - \alpha$

Recall ...

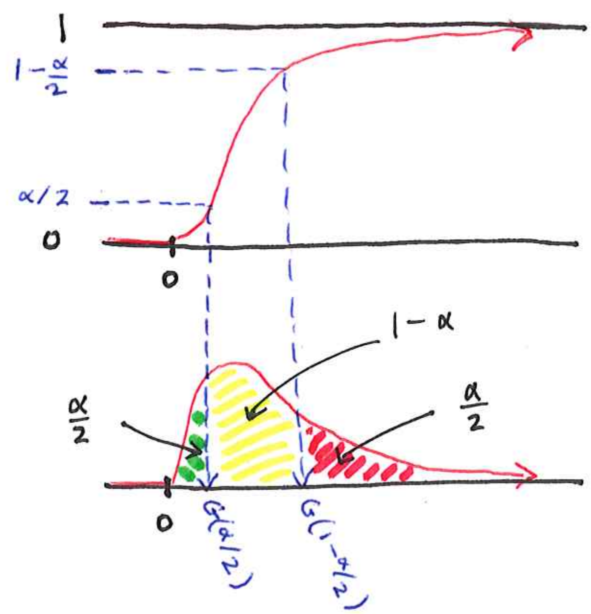
$$X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Gamma}(1, \lambda)$$

$$\sum_{i=1}^{n} X_i \sim \text{Gamma}(n, \lambda)$$

$$\bar{X}_n \sim \text{Gamma}(n, n\lambda)$$

$$\lambda \bar{X}_n \sim \text{Gamma}(n, n)$$

Let $G_{a,b}(u)$ be the quantile function of $\text{Gamma}(a, b)$:



So

$$P\left(G_{n,n}(\alpha/2) < \lambda \bar{X}_n < G_{n,n}\left(1 - \frac{\alpha}{2}\right)\right) = 1 - \alpha$$

$$P\left(\frac{G_{n,n}(\alpha/2)}{\bar{X}_n} < \lambda < \frac{G_{n,n}\left(1 - \frac{\alpha}{2}\right)}{\bar{X}_n}\right) = 1 - \alpha$$

Let's simulate the behavior of the interval . . .

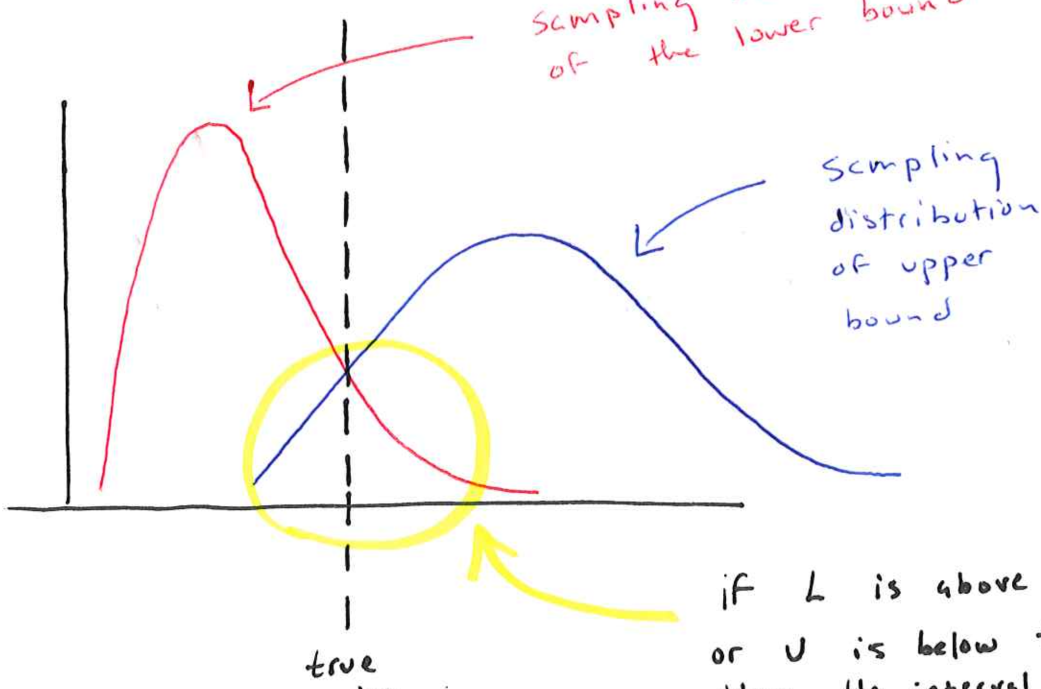
$$\text{Exp}(\lambda = 4)$$

simulate many different datasets of size n from true distribution

X_1	X_1		X_1	X_1
X_2	X_2	repeat	X_2	X_2
$\vdots$	$\vdots$	$M = 5000$ times	$\vdots$	$\vdots$
X_n	X_n		X_n	X_n
$\downarrow$	$\downarrow$	$\cdots$	$\downarrow$	$\downarrow$
(L, U)	(L, U)		(L, U)	(L, U)

each dataset gives a different ~~point estimate~~ 90% confidence interval

90% of these intervals contain the true value. 10% don't.



if L is above true λ or U is below true λ, then the interval fails. that happens $100 \times \alpha\%$ of the time.

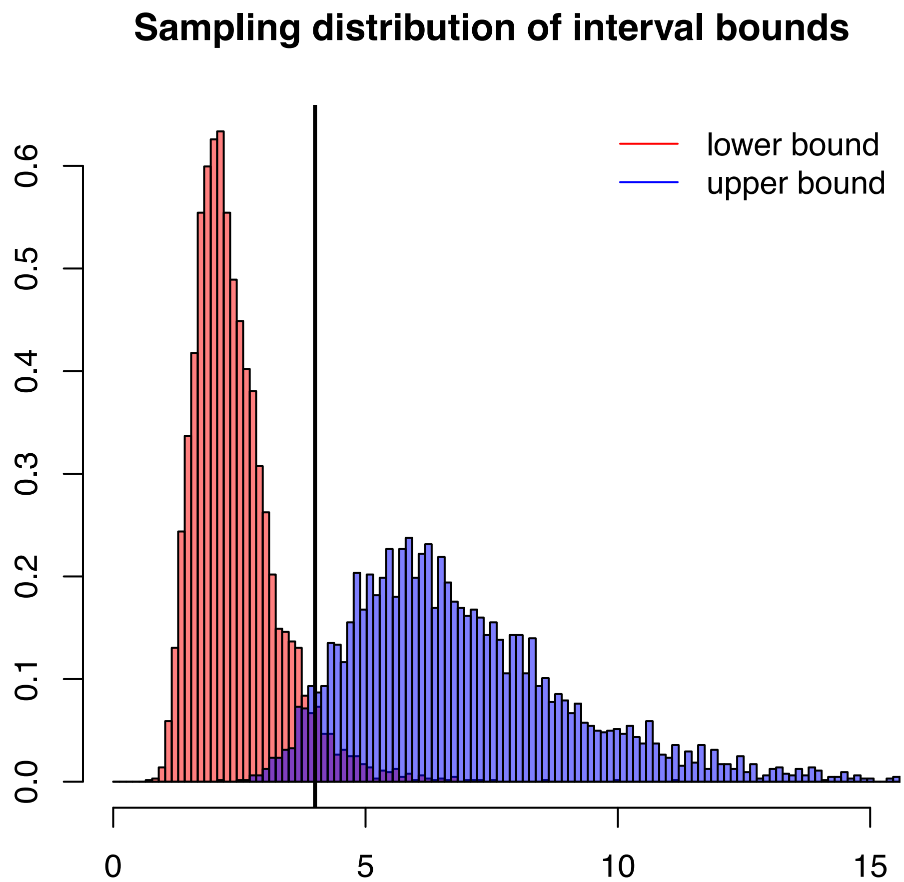
```r
set.seed(12345)

true_lambda <- 4
n <- 10
M <- 5000
truth_in_ci <- numeric(M)
lowers = numeric(M)
uppers = numeric(M)
alpha = 0.1

for(m in 1:M){
  X = rexp(n, true_lambda)
  xbar = mean(X)
  L = qgamma(alpha/2, shape = n, rate = n) / xbar
  U = qgamma(1 - alpha/2, shape = n, rate = n) / xbar
  truth_in_ci[m] = (L < true_lambda) & (true_lambda < U)
  lowers[m] = L
  uppers[m] = U
}

b <- seq(0, max(uppers, lowers), length.out = 250)

hist(lowers, breaks = b, freq = FALSE, xlim = c(0, 15),
     col = rgb(1, 0, 0, alpha = 0.5),
     main = "Sampling distribution of interval bounds",
     xlab = "", ylab = "")
hist(uppers, breaks = b, freq = FALSE, add = TRUE,
     col = rgb(0, 0, 1, alpha = 0.5))
abline(v = true_lambda, lwd = 2)
legend("topright", c("lower bound", "upper bound"), col = c("red", "blue"),
       lty = 1, bty = "n")
```



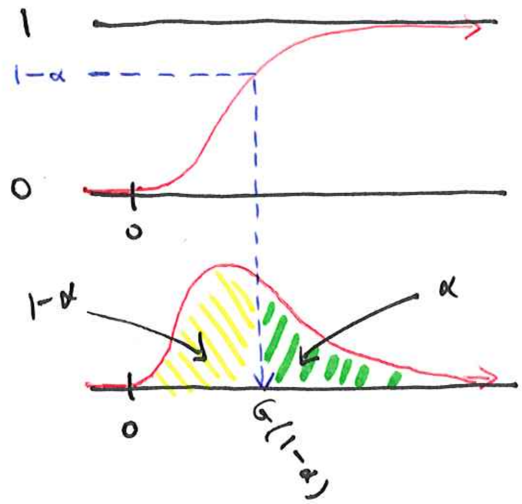
These were 90% confidence intervals, so the fraction of the simulated intervals containing the true value should be close to 0.9:

```
mean(truth_in_ci)
```

```
[1] 0.9018
```

Baller.

The agony and the ecstasy of statistics are that there are a million alternative ways to do everything, so here's another CI w/ the same coverage...



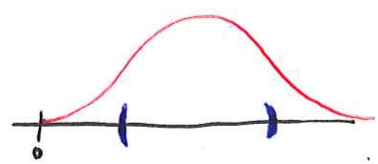
$$P\left(0 < \lambda \bar{X}_n < G_{n,n}(1-\alpha)\right) = 1-\alpha$$

$$P\left(0 < \lambda < \frac{G_{n,n}(1-\alpha)}{\bar{X}_n}\right) = 1-\alpha$$

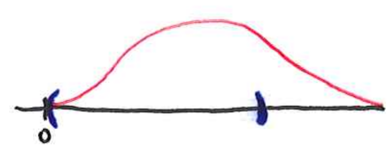
Here's another one... $P(0 < \lambda < \infty) = 1.$

Which do we prefer?

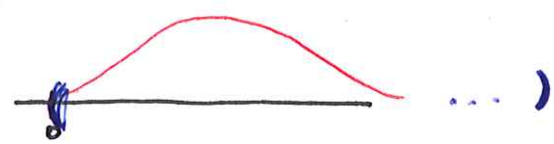
A. $\left(\dfrac{G_{n,n}(\alpha/2)}{\bar{X}_n}, \dfrac{G_{n,n}(1-\frac{\alpha}{2})}{\bar{X}_n}\right)$



B. $\left(0, \dfrac{G_{n,n}(1-\alpha)}{\bar{X}_n}\right)$



C. $(0, \infty)$



Probably A. You want an interval estimate that is wide enough to contain the truth with high confidence, but narrow enough to be informative about where the true parameter lives. Given two competing intervals with the same coverage, go with the smaller one.

MLE in general

pdf or pmf

Data: $X_1, X_2, \ldots, X_n \overset{iid}{\sim} f(x|\theta)$

true but unknown parameter(s)

likelihood function: $L(\theta|X_{1:n}) = f(X_1, X_2, \ldots, X_n|\lambda) = \prod_{i=1}^{n} f(X_i|\theta)$

log-likelihood function:

$$\ell(\theta|X_{1:n}) = \ln L(\theta|X_{1:n})$$
$$= \ln \prod_{i=1}^{n} f(X_i|\theta)$$
$$= \sum_{i=1}^{n} \ln f(X_i|\theta)$$

MLE: $\hat{\theta}_n^{(MLE)} = \underset{\theta}{\operatorname{argmax}}\, L(\theta|X_{1:n}) = \underset{\theta}{\operatorname{argmax}}\, \ell(\theta|X_{1:n})$

properties:

consistent... $\hat{\theta}_n^{(MLE)} \xrightarrow{prob} \theta$

asymptotic normality...

$$\frac{\hat{\theta}_n^{(MLE)} - \theta}{\widehat{se}\left(\hat{\theta}_n^{(MLE)}\right)} \xrightarrow{dist} N(0,1)$$

approximate confidence interval... $\hat{\theta}_n^{(MLE)} \pm z^*\left(1-\frac{\alpha}{2}\right) \widehat{se}\left(\hat{\theta}_n^{(MLE)}\right)$

$P(L_n < \theta < U_n) \approx 1-\alpha$ For large enough n

bias... $E\left(\hat{\theta}_n^{(MLE)}\right) = ???$

may or may not be unbiased. no guarantee in general.

invariance... MLE of $g(\theta)$ is $g\left(\hat{\theta}_n^{(MLE)}\right)$.

STA 3321:

standard error... $se\left(\hat{\theta}_n^{(MLE)}\right) = \sqrt{1 \big/ I_n\left(\hat{\theta}_n^{(MLE)}\right)}$

Fisher information... $I_n\left(\hat{\theta}_n^{(MLE)}\right) = -n \int_{-\infty}^{\infty} f(x|\theta) \frac{\partial^2}{\partial \theta^2} \ln f(x|\theta)\, dx$

Bernoulli

$$X_1, X_2, \dots, X_n \overset{iid}{\sim} \text{Bern}(p) \qquad X_i = \begin{cases} 0 \\ 1 \end{cases}$$

$$L(p \mid X_{1:n}) = \prod_{i=1}^{n} p^{X_i} (1-p)^{1-X_i}$$

$$= p^{\sum_{i=1}^{n} X_i} (1-p)^{n - \sum_{i=1}^{n} X_i}$$

$$\ln L(p \mid X_{1:n}) = \left(\sum_{i=1}^{n} X_i\right) \ln p + \left(n - \sum_{i=1}^{n} X_i\right) \ln(1-p)$$

$$\frac{d \ln L}{dp} = \frac{\sum_{i=1}^{n} X_i}{p} - \frac{\left(n - \sum_{i=1}^{n} X_i\right)}{1-p} = 0$$

$$\frac{\sum_{i=1}^{n} X_i}{p} = \frac{n - \sum_{i=1}^{n} X_i}{1-p}$$

$$\boxed{\hat{p}_n^{(MLE)} = \bar{X}_n}$$

$$(1-p) \sum_{i=1}^{n} X_i = p\left(n - \sum_{i=1}^{n} X_i\right)$$

$$\sum_{i=1}^{n} X_i - \cancel{p \sum_{i=1}^{n} X_i} = pn - \cancel{p \sum_{i=1}^{n} X_i}$$

$$\sum_{i=1}^{n} X_i = pn$$

$$\frac{\sum_{i=1}^{n} X_i}{n} = p$$

Poisson

$$X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Poisson}(\lambda)$$

$$L(\lambda | X_{1:n}) = \prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{X_i}}{X_i!}$$

$$\ln L(\lambda | X_{1:n}) = \sum_{i=1}^{n} \ln\left(e^{-\lambda} \frac{\lambda^{X_i}}{X_i!}\right)$$

$$= \sum_{i=1}^{n} \left[\ln e^{-\lambda} + \ln \lambda^{X_i} - \ln(X_i!)\right]$$

$$= \sum_{i=1}^{n} \left[-\lambda + X_i \ln \lambda - \ln(X_i!)\right]$$

$$= -n\lambda + (\ln \lambda) \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \ln(X_i!)$$

$$\frac{d \ln L}{d\lambda} = -n + \frac{\sum_{i=1}^{n} X_i}{\lambda} = 0$$

$$\frac{\sum_{i=1}^{n} X_i}{\lambda} = n$$

$$\frac{\sum_{i=1}^{n} X_i}{n} = \lambda$$

$$\boxed{\hat{\lambda}_n^{(MLE)} = \bar{X}_n}$$

Normal

$$X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$$

$$L(\mu, \sigma \mid X_{1:n}) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{(X_i - \mu)^2}{\sigma^2}\right)$$

$$= \frac{1}{\sigma^n} \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2\right)$$

$$\ln L(\mu, \sigma \mid X_{1:n}) = -n\ln\sigma - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2$$

Want: $(\hat{\mu}_n, \hat{\sigma}_n) = \underset{\mu,\sigma}{\operatorname{argmax}} \, L(\mu, \sigma \mid X_{1:n})$ multivariable optimization!

$$\frac{\partial \ln L}{\partial \mu} = -\frac{1}{2}\frac{1}{\sigma^2}\sum_{i=1}^{n}(-2)(X_i - \mu) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu) = \frac{1}{\sigma^2}\left[\sum_{i=1}^{n}X_i - \sum_{i=1}^{n}\mu\right] = \frac{1}{\sigma^2}\left[\sum_{i=1}^{n}X_i - n\mu\right]$$

$$\frac{\partial \ln L}{\partial \sigma} = \frac{-n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(X_i - \mu)^2$$

partial derivatives!

Need to solve this system of equations:

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}X_i - \frac{n\mu}{\sigma^2} = 0$$

$$\frac{\partial \ln L}{\partial \sigma} = \frac{-n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(X_i - \mu)^2 = 0$$

And then we have to check the Hessian matrix to make sure our solution isn't a saddle point. Ew!

Solve the first equation . . .

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n} X_i - \frac{n\mu}{\sigma^2} = 0$$

$$\frac{1}{\sigma^2}\sum_{i=1}^{n} X_i = \frac{n\mu}{\sigma^2}$$

$$\sum_{i=1}^{n} X_i = n\mu$$

$$\frac{\sum_{i=1}^{n} X_i}{n} = \mu$$

$$\boxed{\hat{\mu}_n^{(MLE)} = \overline{X}_n}$$

substitute into the second and solve . . .

$$\frac{\partial \ln L}{\partial \sigma} = \frac{-n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n} (X_i-\mu)^2 = 0$$

$$\frac{1}{\sigma^3}\sum_{i=1}^{n} (X_i-\mu)^2 = \frac{n}{\sigma}$$

$$\sum_{i=1}^{n} (X_i-\mu)^2 = n\sigma^2$$

$$\frac{1}{n}\sum_{i=1}^{n} (X_i-\mu)^2 = \sigma^2$$

$$\boxed{\widehat{\sigma^2_n}^{(MLE)} = \frac{1}{n}\sum_{i=1}^{n} (X_i-\overline{X}_n)^2}$$

This is biased! Our usual estimator for the variance is $\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2$, which is unbiased. So, MLE gives an answer, but not necessarily the "best" answer.

From make-up lecture

$$X_1, X_2, \ldots, X_n \stackrel{iid}{\sim} N(\mu, \sigma^2) \implies \bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$\implies \widehat{\sigma_n^2}^{(MLE)} \sim \text{Gamma}\left(\frac{n-1}{2}, \frac{n}{2}\frac{1}{\sigma^2}\right)$$

$$\implies \frac{\widehat{\sigma_n^2}^{(MLE)}}{\sigma^2} \sim \text{Gamma}\left(\frac{n-1}{2}, \frac{n}{2}\right)$$

$$\implies \frac{\bar{X}_n - \mu}{\sqrt{\frac{1}{n-1}\widehat{\sigma_n^2}^{(MLE)}}} \sim t_{n-1}$$

Exact confidence intervals . . .

<u>Mean</u>: $\bar{X}_n \pm t_{n-1}\left(1-\frac{\alpha}{2}\right)\sqrt{\frac{1}{n-1}\widehat{\sigma_n^2}^{(MLE)}}$

<u>Variance</u>:

$$\underbrace{\left(\frac{\widehat{\sigma_n^2}^{(MLE)}}{G_{\frac{n-1}{2},\frac{n}{2}}\left(1-\frac{\alpha}{2}\right)}, \frac{\widehat{\sigma_n^2}^{(MLE)}}{G_{\frac{n-1}{2},\frac{n}{2}}\left(\frac{\alpha}{2}\right)}\right)}$$

$$P\left(G_{\frac{n-1}{2},\frac{n}{2}}(\alpha/2) < \frac{\widehat{\sigma_n^2}^{(MLE)}}{\sigma^2} < G_{\frac{n-1}{2},\frac{n}{2}}\left(1-\frac{\alpha}{2}\right)\right) = 1-\alpha$$

$$\implies P\left(\frac{\widehat{\sigma_n^2}^{(MLE)}}{G_{\frac{n-1}{2},\frac{n}{2}}\left(1-\frac{\alpha}{2}\right)} < \sigma^2 < \frac{\widehat{\sigma_n^2}^{(MLE)}}{G_{\frac{n-1}{2},\frac{n}{2}}(\alpha/2)}\right) = 1-\alpha$$