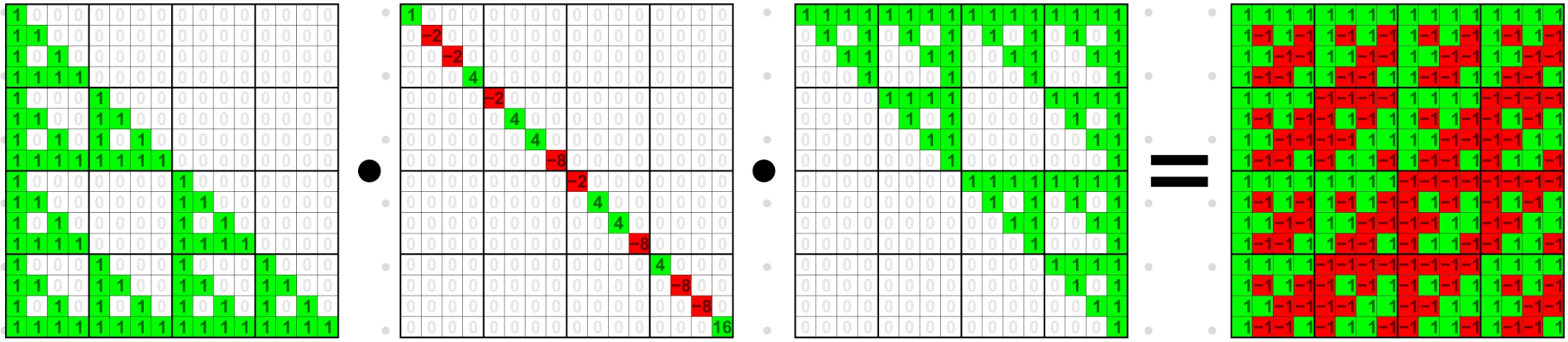


Statistics

Lecture 8

a general introduction to inference.

MATH 414

Exercises for assignment

① m.l.e. for binomial, uniform distributions etc.

Confidence Intervals

Defn Let $X_1, X_2, \ldots, X_n$ be a sample on a random variable X, where X has p.d.f. $f(x;\theta)$, $\theta \in \Omega$.

Let $0 < \alpha < 1$ be specified. Let $L = L(X_1, \ldots, X_n)$ and $U = U(X_1, \ldots, X_n)$ be two statistics. We say that the interval (L, U) is $(1-\alpha)100\%$ confidence interval for θ if

$$1 - \alpha = P_\theta(\theta \in (L, U)).$$

Once the sample is drawn, the realized value of the confidence interval is (l, u), an interval of real numbers. Therefore, θ either belongs to (l, u) or it does not.

Once the sample is drawn, the realized value of the confidence interval is (l, u), an interval of real numbers. Therefore, θ either belongs to (l, u) or it does not.

Therefore, we could think of confidence interval interms of Bernoulli trials with $p = 1 - \alpha$.

Once the sample is drawn, the realized value of the confidence interval is (l, u), an interval of real numbers. Therefore, θ either belongs to (l, u) or it does not.

Therefore, we could think of confidence interval interms of Bernoulli trials with $p = 1-\alpha$.

If one makes M $(1-\alpha)100\%$ confidence intervals then we expect that $(1-\alpha) \cdot M$ of them will contain θ..

Example Confidence interval for μ under normality

Suppose $X_1, X_2, \ldots, X_n$ are $N(\mu, \sigma^2)$ distributed.

Let $\bar{X}$ and S^2 denote the sample mean and sample variance respectively.

We proved that $T \equiv (\bar{X}-\mu)/(S/\sqrt{n})$ has a t-distribution with n degrees of freedom.

Example Confidence interval for μ under normality

Suppose $X_1, X_2, \ldots, X_n$ are $N(\mu, \sigma^2)$ distributed.

Let $\bar{X}$ and S^2 denote the sample mean and sample variance respectively.

We proved that $T \equiv (\bar{X}-\mu)/(S/\sqrt{n})$ has a t-distribution with n degrees of freedom.

The random variable T is called our pivot r.v.

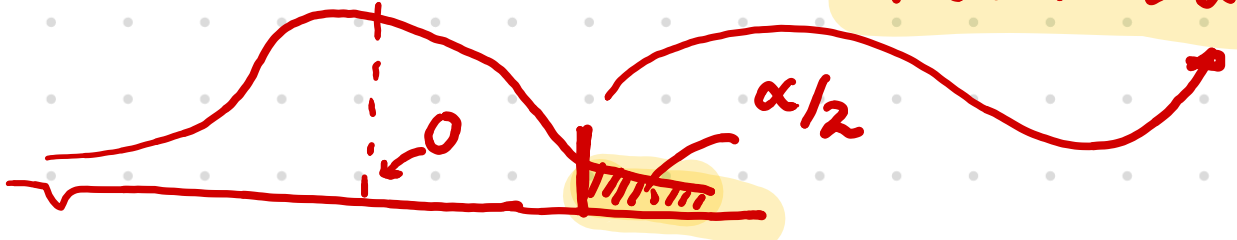
Example — Confidence interval for μ under normality

Suppose $X_1, X_2, \ldots, X_n$ are $N(\mu, \sigma^2)$ distributed.

Let $\bar{X}$ and S^2 denote the sample mean and sample variance respectively.

We proved that $T \equiv (\bar{X}-\mu)/(S/\sqrt{n})$ has a t-distribution with n d rees of freedom.

For $0 < \alpha < 1$, define $t_{\alpha/2, n-1}$ to be the value of T such that $P(T > t_{\alpha/2, n-1}) = \alpha/2$.



$$\therefore \quad 1-\alpha = P\left(-t_{\alpha/2,n-1} < T < t_{\alpha/2,n-1}\right)$$

$$= P\left(-t_{\alpha/2,n-1} < \frac{\bar{X}-\mu}{S/\sqrt{n}} < t_{\alpha/2,n-1}\right)$$

$$= P\left(-t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} < \bar{X}-\mu < t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right)$$

$$= P\left(\bar{X}-t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} < \mu < \bar{X}+t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right)$$

$$\therefore \quad 1-\alpha = P\left(-t_{\alpha/2,n-1} < T < t_{\alpha/2,n-1}\right)$$

$$= P\left(-t_{\alpha/2,n-1} < \frac{\bar{X}-\mu}{S/\sqrt{n}} < t_{\alpha/2,n-1}\right)$$

$$= P\left(-t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} < \bar{X}-\mu < t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right)$$

$$= P\left(\bar{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right)$$

Once the sample is drawn, let $\bar{x}$ and s denote the sample mean and sample variance.

Then a $(1-\alpha)100\%$ confidence interval of μ is given by $\left(\bar{x} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}},\ \bar{x} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right)$

Example.

nhtemp is a dataset in R

```
t.test(nhtemp, conf.level = 0.99)
```

Example.

nhtemp is a dataset in R

t.test(nhtemp, conf.level = 0.99)

Confidence Interval for Proportion

of $X \sim Ber(1, p)$ and $X_1, X_2, \ldots, X_n$ is a random sample. Then $\bar{X}$ is the proportion of successes. With the large sample assumption, this comes to

$$\left(\hat{p} \pm z_{\alpha/2} \sqrt{\hat{p}(1-\hat{p})/n}\right)$$

Example.

nhtemp is a dataset in R

t.test(nhtemp, conf.level = 0.99)

Confidence Interval for Proportion of $X \sim Ber(1, p)$ and $X_1, X_2, \ldots, X_n$ is a random sample. Then $\bar{X}$ is the proportion of successes. With the large sample assumption, this comes to

$$\left(\hat{p} \pm z_{\alpha/2} \sqrt{\hat{p}(1-\hat{p})/n}\right)$$

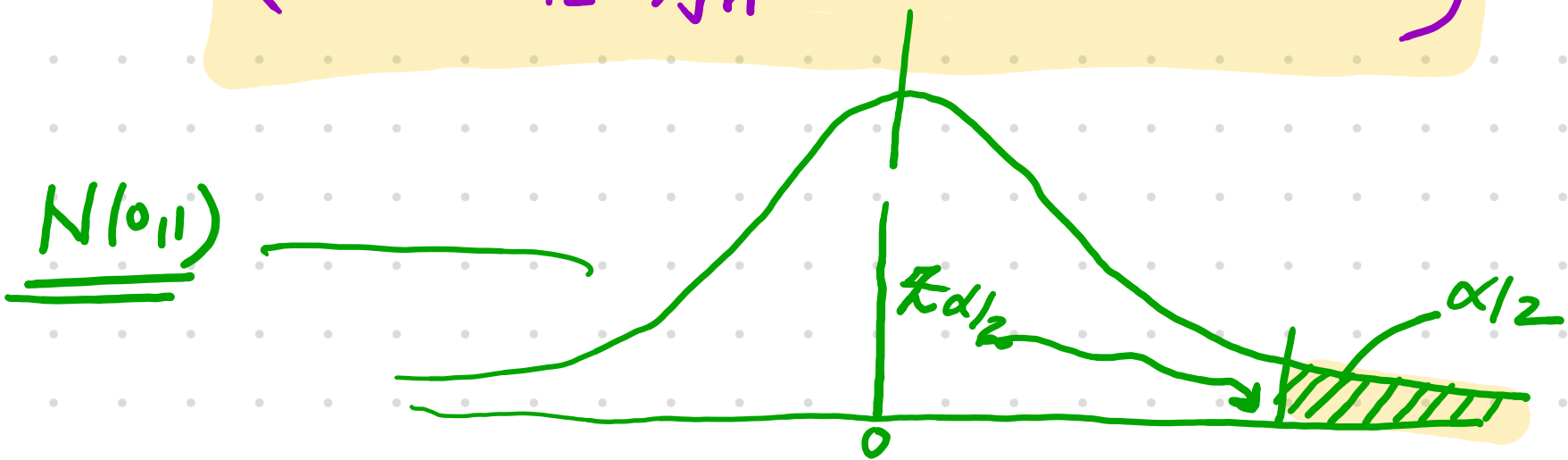
If there are 28 successes in 143, we

prop.test(28, 143, conf.level = 0.99)

This is based on the normality of the distribution.

If normality is not known but sample is large, we may use CLT to find approximate confidence intervals.

$$\left(\bar{x} - z_{\alpha/2}\, s/\sqrt{n},\ \bar{x} + z_{\alpha/2}\, s/\sqrt{n}\right)$$



Confidence Intervals for Difference in Means

How to compare means of X and Y?

Let the means be μ_1 and μ_2 respectively.

We will obtain confidence intervals for the difference $\Delta = \mu_1 - \mu_2$. Assume that σ_1^2 and σ_2^2 are variances of X and Y respectively. Let $X_1, X_2, \ldots, X_n$ be a random sample from X and $Y_1, Y_2, \ldots, Y_n$ be a random sample from Y.

$$\bar{X} = \frac{\sum X_i}{n}, \quad \bar{Y} = \frac{\sum Y_i}{n}, \quad \hat{\Delta} = \bar{X} - \bar{Y}.$$

$\hat{\Delta}$ is an unbiased estimator.

The difference $\hat{\Delta} - \Delta$ will be the numerator of the pivot variable.

By independence, $\text{Var}(\hat{\Delta}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$

Let $S_1^2 = \frac{\sum_{i=1}^{n_1} (X_i - \bar{X})^2}{n_1 - 1}$ and $S_2 = \frac{\sum_{i=1}^{n_2} (Y_i - \bar{Y})^2}{n_2 - 1}$

We can consider the pivot variable

$$Z = \frac{\hat{\Delta} - \Delta}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

whence an approximate confidence interval is given by

$$\left((\bar{x} - \bar{y}) - z_{\alpha/2} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \, , \, (\bar{x} - \bar{y}) + z_{\alpha/2} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \right).$$

As an example, let us compute an exact confidence interval when X and Y are normal with the same variance, $\sigma_1^2 = \sigma_2^2$.

As an example, let us compute an exact confidence interval when X and Y are normal with the same variance, $\sigma_1^2 = \sigma_2^2$. Let $X \sim N(\mu_1, \sigma^2)$ and $Y \sim N(\mu_2, \sigma^2)$.

Note that

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

has $N(0,1)$ distribution.

This will be the numerator of our pivot variable.

Consider

$$S_p = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

which is a weighted average of S_1 and S_2.

Consider

$$S_p = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

which is a weighted average of S_1 and S_2. (Check that this is an unbiased estimator for σ^2).

Consider

$$S_p = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

which is a weighted average of S_1 and S_2. (Check that this is an unbiased estimator for σ^2).

Now, $\frac{(n_1-1)S_1^2}{\sigma^2}$ has a $\chi^2(n_1-1)$ distribution

$\frac{(n_2-1)S_2^2}{\sigma^2}$ has a $\chi^2(n_2-1)$ distribution

Consider

$$S_p = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

which is a weighted average of S_1 and S_2. (Check that this is an unbiased estimator for σ^2).

Now, these are independent:

$$\frac{(n_1-1)S_1^2}{\sigma^2} \text{ has a } \chi^2(n_1-1) \text{ distribution}$$

$$\frac{(n_2-1)S_2^2}{\sigma^2} \text{ has a } \chi^2(n_2-1) \text{ distribution}$$

Consider

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

which is a weighted average of s_1 and s_2. (Check that this is an unbiased estimator for σ^2).

Now, these are independent:

$$\frac{(n_1-1)s_1^2}{\sigma^2} \text{ has a } \chi^2(n_1-1) \text{ distribution}$$

$$\frac{(n_2-1)s_2^2}{\sigma^2} \text{ has a } \chi^2(n_2-1) \text{ distribution}$$

$\therefore \frac{(n-2)s_p^2}{\sigma^2}$ has a $\chi^2(n-2)$ distribution.

Due to the independence of S_1^2 and $\bar{X}$, S_2^2 and $\bar{Y}$, and the independence of the samples, S_p^2 is independent of $\bar{X}-\bar{Y}$.

Due to the independence of S_1^2 and $\bar{X}$, S_2^2 and $\bar{Y}$, and the independence of the samples, S_p^2 is independent of $\bar{X}-\bar{Y}$. Therefore, the statistic

$$T = \frac{\left[(\bar{X}-\bar{Y}) - (\mu_1-\mu_2)\right] / \sigma\sqrt{1/n_1 + 1/n_2}}{\sqrt{(n-2)S_p^2/(n-2)\sigma^2}}$$

$$= \frac{(\bar{X}-\bar{Y}) - (\mu_1-\mu_2)}{\sqrt{S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \sim t_{n-2}$$

This gives us a confidence interval

$$\left((\bar{x}-\bar{y}) - t_{\alpha/2,n-2}\, s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \;,\; (\bar{x}-\bar{y}) + t_{\alpha/2,n-2}\, s_p \sqrt{\frac{1}{n}+\frac{1}{n_2}} \right)$$

This gives us a confidence interval

$$\left((\bar{x}-\bar{y}) - t_{\alpha/2,n-2}\, s_p \sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \;,\; (\bar{x}-\bar{y}) + t_{\alpha/2,n-2}\, s_p \sqrt{\frac{1}{n}+\frac{1}{n_2}} \right)$$

```
install.packages("remotes")
remotes::install_github("joemckean/mathstat",
                        force = TRUE)

require(mathstat)
hitht = bb$height[bb$hitpitind == 1]
pitht = bb$height[bb$hitpitind == 0]
t.test(pitht, hitht, var.equal = T)
```

Confidence Interval for Variance.

Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$ where both are unknown.

$\frac{(n-1)S^2}{\sigma^2}$ is a random variable with a $\chi^2(n-1)$ distribution.

Find 'b' so that $P((n-1)S^2/\sigma^2 < b) = 0.975$

$b = qchisq(0.975, n-1)$

Confidence Interval for Variance.

Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$ where both are unknown.

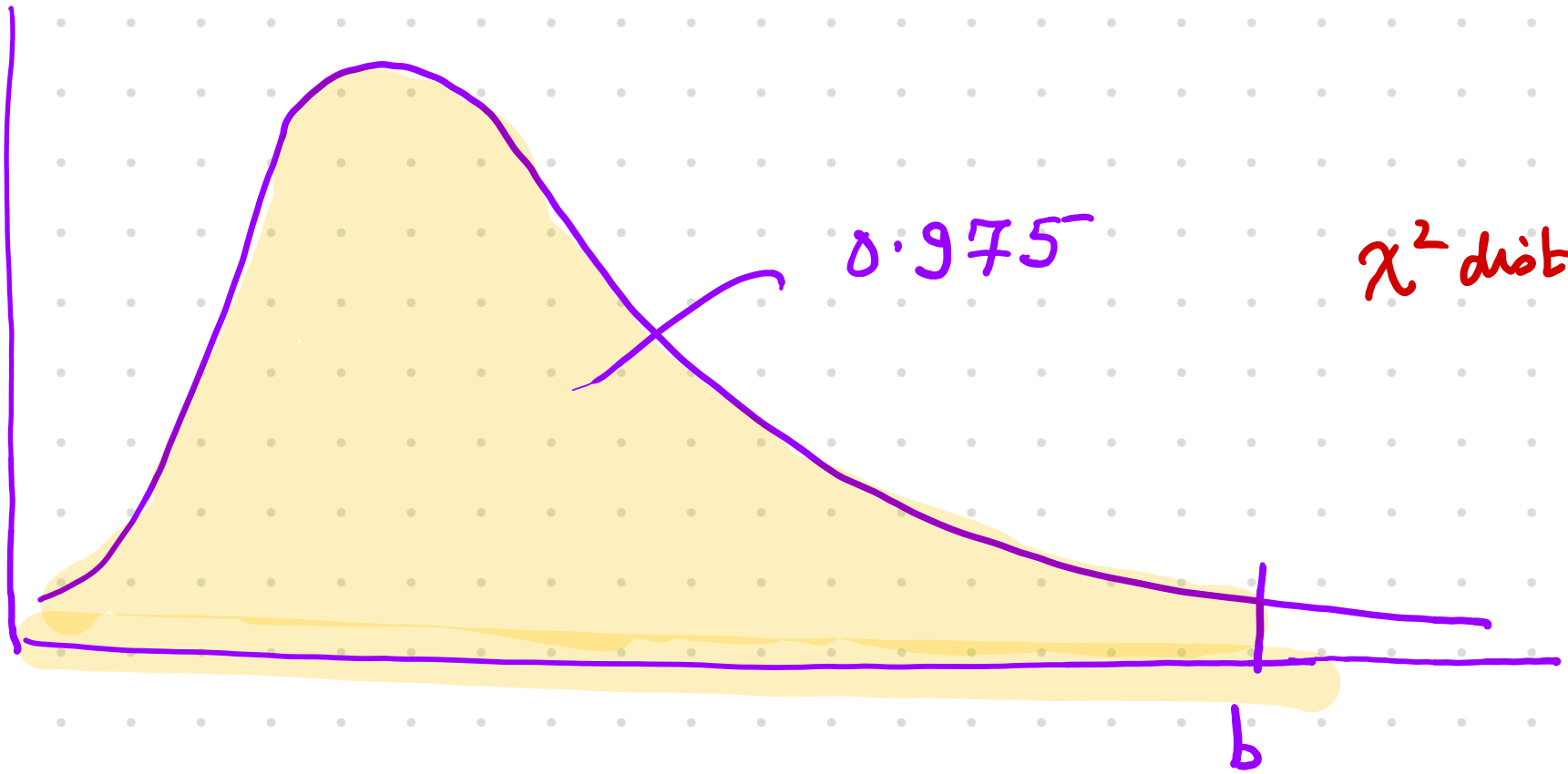
$\frac{(n-1)S^2}{\sigma^2}$ is a random variable with a $\chi^2(n-1)$ distribution.

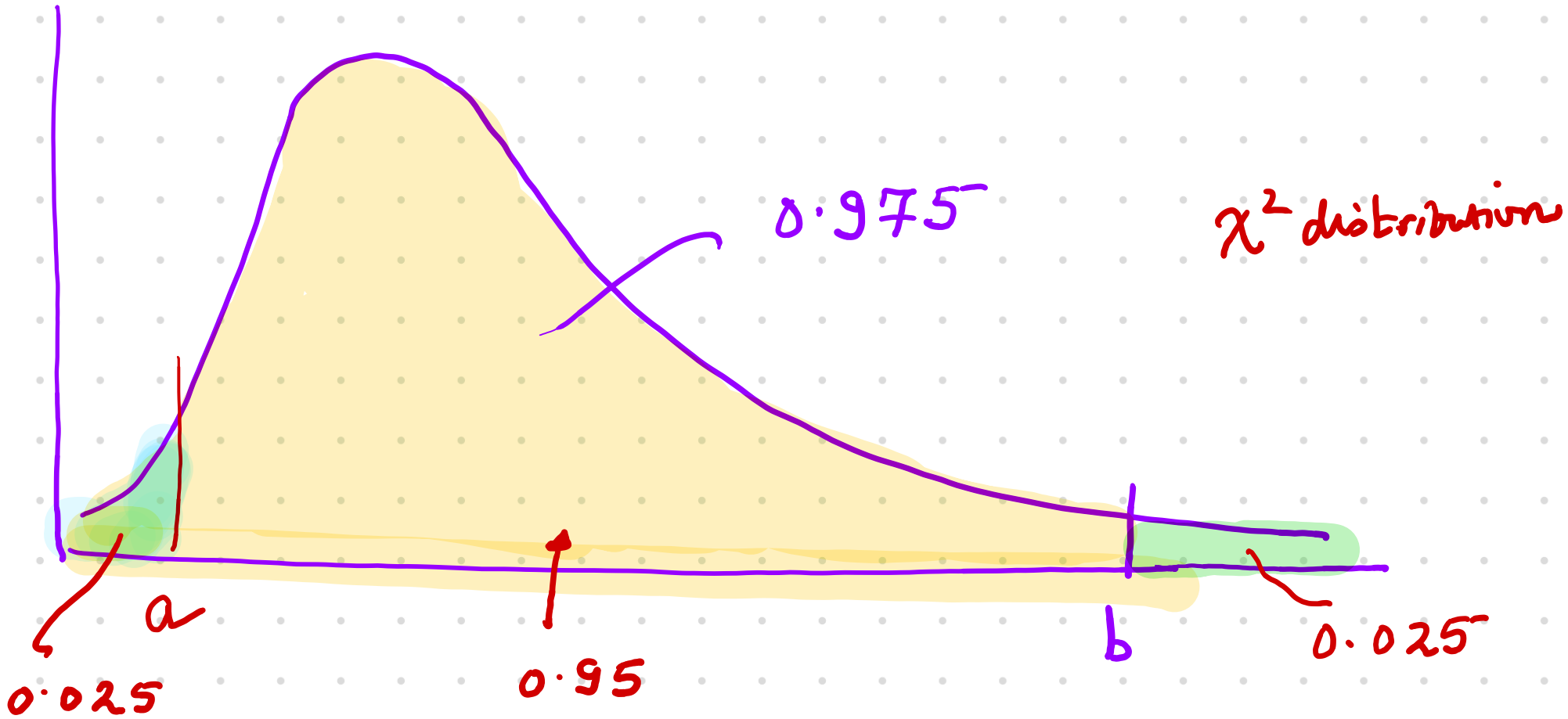
Find 'b' so that $P((n-1)S^2/\sigma^2 < b) = 0.975$

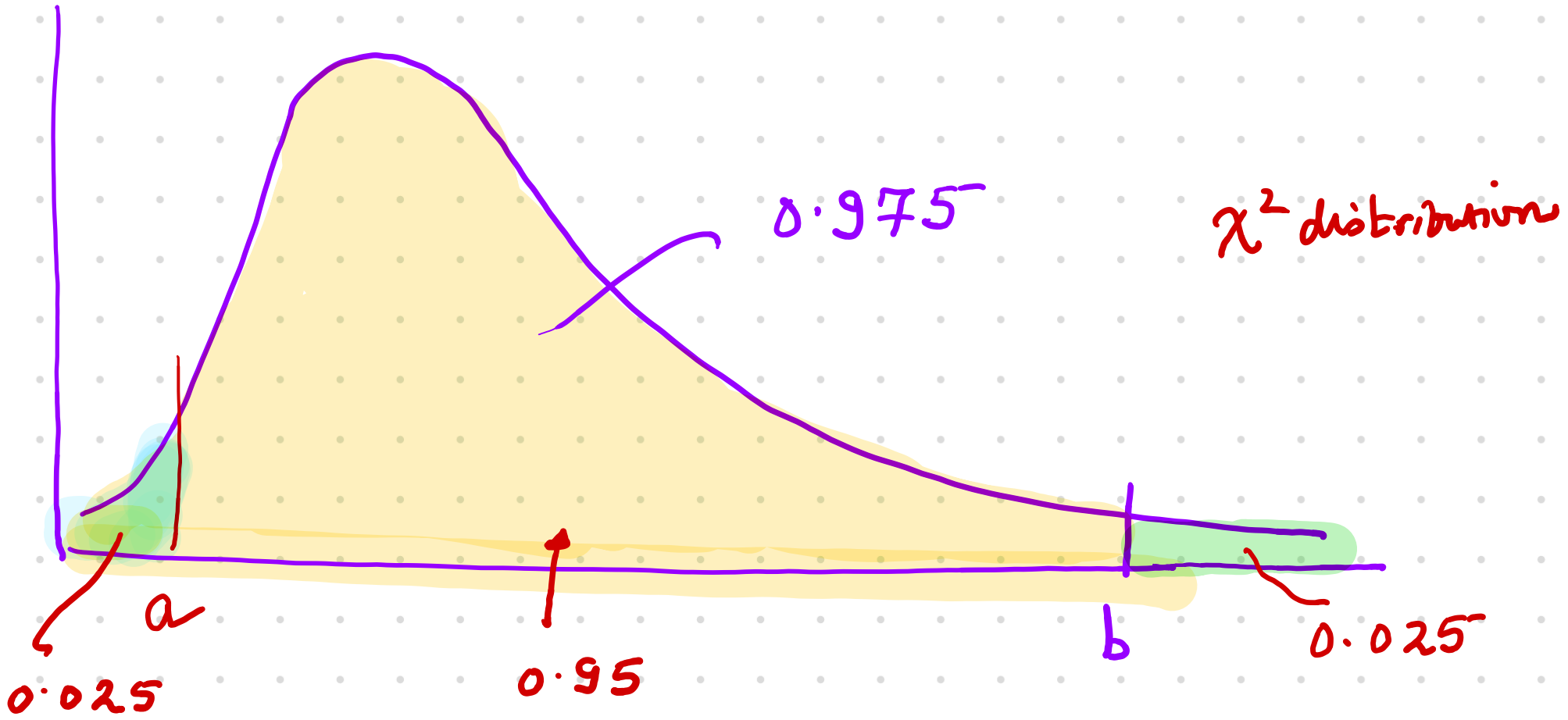
$b = qchisq(0.975, n-1)$

Find 'a' so that $P(a < \frac{(n-1)S^2}{\sigma^2} < b) = 0.95$

$a = qchisq(0.025, n-1)$


0.975
χ^2 distribution
b


0.975
χ^2 distribution
a
b
0.025
0.95
0.025



This is the same as

$$P\left(\frac{(n-1)s^2}{b} < \sigma^2 < \frac{(n-1)s^2}{a}\right) = 0.95$$

Q. If $n = 9$ and $s^2 = 7.93$, find a 95% confidence interval for σ^2.