Methods of Mathematical Statistics

Notes by Tim Brown and Guoqi Qian

Module 5: Distributions of Functions of RVs

Contents

1 Transformations of One Random Variable — 5.1, 5.4

1.1 Examples using CDF method

Two methods — MGF and CDF

MGF method If $Z \sim N(0, 1)$ (recall that $\sim$ means "(is) distributed as"), the MGF of $X = Z^2$ was found by the "make the answer into a probability density" procedure and the MGF was that of $\chi^2(1)$ (see p.28 of Module 3)

Because MGF's determine the distribution this showed that the distribution of Z^2 is $\chi^2(1)$

CDF method For $U \sim U(0,1)$, we found the cdf and then the pdf of $2U+2$ in the "Other Random Numbers" example (pp.20 and 21 of Module 3)

Essential step was to express the event $[2U+2 \leq y]$ as $[U \leq \frac{y-2}{2}]$ (after determining the relevant values of y)

CDF & PDF for χ^2

Suppose $Z \sim N(0,1)$. Let Φ & ϕ be the CDF & PDF of Z.
Then, for any sample outcome s and non-negative number x, $X = Z^2(s) \leq x$ if, and only if, $-\sqrt{x} \leq Z(s) \leq \sqrt{x}$. That is, the events satisfy $[X \leq x] = [-\sqrt{x} \leq Z \leq \sqrt{x}]$. Hence, their probabilities are the same:

$$P(X \leq x) = P(-\sqrt{x} \leq Z \leq \sqrt{x}).$$

Writing in terms of CDF's, F_X, Φ: $F_X(x) = \Phi(\sqrt{x}) - \Phi(-\sqrt{x})$. Differentiating using the chain rule gives the PDF of X:

$$\begin{aligned} f_{X^2}(x) &= \frac{1}{2\sqrt{x}}(\phi(\sqrt{x}) + \phi(-\sqrt{x})) \\ &= \frac{2}{2\sqrt{x}}\phi(\sqrt{x}) \\ &= \frac{1}{\sqrt{2\pi}\sqrt{x}}e^{-\frac{x}{2}}. \end{aligned}$$

1.2 Generating random numbers

Example — CDF Method

If $U \sim U(0,1)$ and F is *any* cdf with π_p denoting the pth quantile of F, then $X = \pi_U$ has distribution F (see next slide where the calculations are an extension of the "Other Random Numbers" example)

This is the reason that it is only necessary to generate uniform random numbers on a computer, because any other distribution can, in principle, be obtained by applying its quantile function to the uniform random numbers

In practice the quantile function can be difficult to compute, so it is not used directly in many cases (but sometimes indirectly as for normal random variables — Box-Muller Transformation textbook p. 186)

Generating Random Numbers from continuous distributions

Suppose X has continuous distribution function F which is strictly increasing. Then, the quantile function, π_p, is the inverse of F. That is, the events $[X = \pi_U \leq x] = [U \leq F(x)]$. Hence, their probabilities are the same:

$$P(X = \pi_U \leq x) = P(U \leq F(x)).$$

Writing in terms of CDF's, F_X: $F_X(x) = F_U(F(x)) = F(x)$, since the distribution function of U is the identity on $[0,1]$ and $0 \leq F(x) \leq 1$. The argument is the same for more general distribution functions, for example discrete ones, but we need to be careful in the definition of the quantile function to make the events equal.

2 Transformations of Two Random Variables — 5.2

2.1 Summary

Two methods, MGF & CDF

Textbook considers the analogue to the CDF method in a number of cases (pp. 179 to 187).

Maximum and Minimum of independent uniform random variables on pp.4-6 of Module 4 was an example of these techniques.

MGF method not used in textbook.

Two variable transformations of rv's not covered further in MAST90105 except results will be used for particular statistical problems — see Sample Variance below.

3 Several Random Variables — 5.3

3.1 Summary

Expectation and Variance

In Module 2 and Module 4, vital results about expectation and variance of a *sum* of random variables were discussed.

Textbook has these on p.192.

4 Random Variables associated with the Normal Distribution - 5.5

4.1 Sums of constants times normal rv's

Linear combination of normal rv's is normal

If $X_1 \sim N(\mu_1, \sigma_1^2), \ldots, X_n \sim N(\mu_n, \sigma_n^2)$ are independent random variables, then the sum

$$\sum_{i=1}^{n} a_i X_i \sim N(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2)$$

Proof is an extension (by induction) of the one that we did in Module 4 p. 12.

In particular, putting $a_i = \frac{1}{n}, i = 1, \cdots, n$, the sample mean, $\frac{\sum_{i=1}^{n} X_i}{n}$, for independent *normal* random variables is *exactly normally distributed* with expected value equal to the population mean μ and variance equal to $\frac{\sigma^2}{n}$ where σ is the population standard deviation.

The distribution of the sample variance can also be calculated as follows.

4.2 Sample Mean & Variance for Normal RV's

Sample Variance Definition

The sample variance S^2 for random variables $X_1, X_2, \ldots, X_n$ is defined by

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \tag{1}$$

where $\bar{X}$ is the sample mean:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{2}$$

Key Facts

If $X_1, X_2, \cdots, X_n \sim N(\mu, \sigma^2)$ independently of each other, then

1. $\bar{X}$ and S^2 are *independent*,
2. $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$
3. $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

Sample mean discussed above.

Proof of independence deferred to Ch 6 in the textbook — this will not be discussed further in MAST90105.

Given independence of the sample mean and variance, the distribution of the sample variance follows from calculating its mgf.

Key to this calculation is the Analysis of Variance identity. For any numbers $x_1, \cdots, x_n$:

$$\sum_{i=1}^{n}(x_i - \mu)^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 + n(\bar{x} - \mu)^2. \tag{3}$$

Interpretation of AoV identity and degrees of freedom

For data $x_1, \ldots x_n$ regarding the population mean μ as fixed (but unknown), the vector $(x_1 - \mu, \ldots, x_n - \mu)$ are the *model* departures of the data from the population mean.

This model vector can be decomposed:

$$\text{model} = \text{estimate} + \text{residual}$$

where the *estimate* vector is $(\bar{x} - \mu, \cdots, \bar{x} - \mu)$ and the *residual* vector is $(x_1 - \bar{x}, \cdots, x_n - \bar{x})$.

The model vector has n items free to vary (from sample to sample) so has n degrees of freedom.

The estimate vector has only 1 quantity, $\bar{x}$, which is free to vary so has 1 degree of freedom.

The residuals have to add to 0 so one residual is known from the other $n-1$ — the residuals have $n-1$ degrees of freedom.

The Analaysis of Variance identity (see also Module 2, p.22) says that the squared distances add like Pythagoras' Theorem.

Proof of AoV identity

The Analysis of Variance equation (3) is derived as follows:

$$\begin{aligned}\sum_{i=1}^{n}(x_i-\mu)^2 &= \sum_{i=1}^{n}(x_i-\bar{x}+\bar{x}-\mu)^2 \\ &= \sum_{i=1}^{n}((x_i-\bar{x})^2+2(x_i-\bar{x})(\bar{x}-\mu)+(\bar{x}-\mu)^2) \\ &= \sum_{i=1}^{n}(x_i-\bar{x})^2+n(\bar{x}-\mu)^2+2(\bar{x}-\mu)\sum_{i=1}^{n}(x_i-\bar{x}) \\ &= \sum_{i=1}^{n}(x_i-\bar{x})^2+n(\bar{x}-\mu)^2\end{aligned}$$

since the fact that the sample mean is the sum of the rv's divided by n implies that the cross products add to zero.

Key Facts Proof

Divide both sides of the Analysis of Vairance Identity by σ^2, apply the identity to the rv's $X_1, \cdots X_n$ and use independence of the sample mean and variance:

$$E\left(\exp\left[t\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2\right]\right) = E\left(\exp\left[t\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2}\right]\right) \times E\left(\exp\left[t\left(\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}\right)^2\right]\right).$$

So using the definition in equation (1)

$$E\left(\exp\left[t\frac{(n-1)S^2}{\sigma^2}\right]\right) = \frac{E\left(\exp\left[t\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2\right]\right)}{E\left(\exp\left[t\left(\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}\right)^2\right]\right)}. \tag{4}$$

Key Facts Proof 2

The distribution of $\left(\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}\right)^2$ is that of the square of a standard normal random variable so is $\chi^2(1)$.

The distribution of $\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2$ is $\chi^2(n)$ because the terms in the sum are squares of independent standard normal random variables — see the MGF calculations on p.29 of Module 3.

Hence, for any t, from (4)

$$E\left(\exp\left[t\frac{(n-1)S^2}{\sigma}\right]\right) = \frac{(1-2t)^{1/2}}{(1-2t)^{n/2}}$$
$$= \frac{1}{(1-2t)^{(n-1)/2}} \qquad (5)$$

which is the MGF of $\chi^2(n-1)$ as required.

Notes on Key Facts

The slide on degrees of freedom explains the reason that this term is used for this parameter in the χ^2 distribution.

Note that the expectation and variance of a $\chi^2(n-1)$ random variable are $n-1$ and $2(n-1)$ (see Module 3, p.14).

Hence

$$E(S^2) = \sigma^2, \quad Var(S^2) = \frac{2\sigma^4}{n-1}. \qquad (6)$$

Equation (6) is the reason why the Sample Variance definition has $n-1$ rather n in the denominator.

If the Sample Variance had n on the bottom line rather than $n-1$, its expected value would differ from σ^2 in a systematic way depending on the sample size — it would underestimate σ^2 by a factor of $\frac{n-1}{n}$.

For large n, in repeated samples the sample variances will be clustered tightly around the population variance.

Example — Sample Variance

If 5 observations are independently taken of heights of people from a population with mean 170 cm, sd 4 cm and a normal distribution of heights, what is

1. the expected value and 95th percentile of the sample variance,
2. the expected value and 95th percentile of the average of squared departures from the population mean?

Solution — Sample Variance

Suppose that $X_1, X_2, \cdots X_5$ are the heights (in cm) of the randomly chosen people from the population so that $X_i \sim N(170, 16), i = 1, 2, \ldots, 5$ independently of one another.

The sample variance is given by equation (1) with $n = 5$ and the average of squared deviations is $\hat{\sigma}^2 = \frac{\sum_1^5 (X_i - 170)^2}{5}$.

The distributions of S^2 and $\hat{\sigma}^2$ are those of $\chi^2(4)$ & $\chi^2(5)$ random variables divided by their degrees of freedom and multiplied by 16, the population variance.

Expected values of both are thus 16, the population variance.

The 95th percentiles for χ^2 are 11.0705 and 9.487729 from the R commands on the next slide.

The answers are the χ^2 percentiles divided by the df and times 16 ie 37.95 and 35.43

Lower for $\hat{\sigma}^2$ from the extra information in the population mean.

```
(q5 = qchisq(0.95, 5))

## [1] 11.0705

(q4 = qchisq(0.95, 4))

## [1] 9.487729

q4/4 * 16

## [1] 37.95092

q5/5 * 16

## [1] 35.42559
```

4.3 t Distribution

In practice

Data is gathered and assumed to be a random sample from a normal population.

In other words, the data are assumed to be observations from independent and identically distributed normal random variables.

Aim: to learn about the population mean and the population standard deviation from the data.

Sensible to use sample mean as estimate of population mean — will discuss this more in Module 6.

Likewise sensible to use the sample variance, whose expected value is the population variance, as an estimate of the population variance.

From (2)

$$\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \sim N(0,1)$$

What happens if σ is replaced by S?

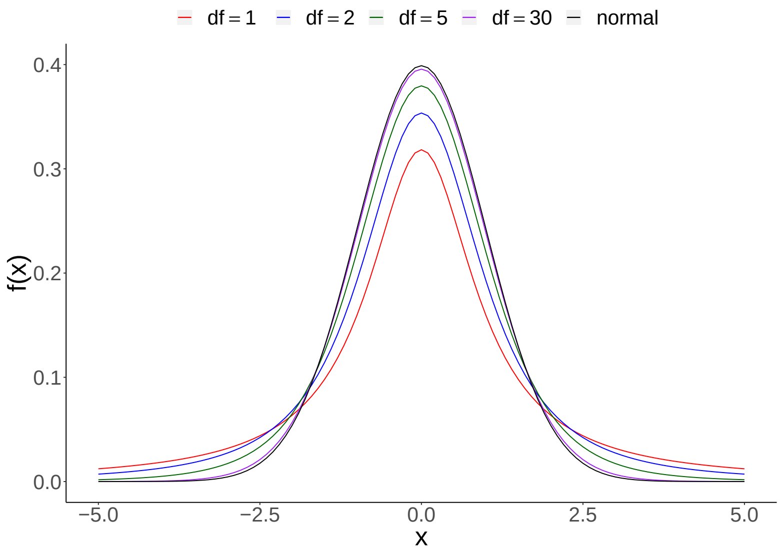

Figure 1: Densities for t distributions for varying degrees of freedom (df)

t Distribution

With r degrees of freedom is defined as that of the rv

$$T = \frac{\sqrt{r}Z}{\sqrt{R}} \tag{7}$$

where $Z \sim N(0,1)$, $R \sim \chi^2(r)$ and Z, R are independent.

This is written as $T \sim t(r)$.

Percentiles of the t-distribution give are necessary in estimating margins of error in confidence intervals (see Module 7), assuming that observations are normally distributed.

T statistic from a normal sample

Suppose $X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$ independently of each other, the sample mean, $\bar{X}$, is defined by equation (2) and the sample variance, S^2, is defined by equation (1).

Then

$$T = \frac{\sqrt{n}(\bar{X} - \mu)}{S} \sim t_{n-1} \tag{8}$$

Because

$$T = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \times \sqrt{n-1}\sqrt{\frac{\sigma^2}{(n-1)S^2}}$$

and the two terms on the right are independent by Key Fact (1), the first term has N(0,1) distribution by Key Fact (2) and the second term is of the form $\frac{\sqrt{n-1}}{\sqrt{R}}$ where $R \sim \chi^2(n-1)$ by Key Fact (3).

T statistic from normal sample Ctd

Note The quantity $\frac{\sigma}{\sqrt{n}}$ is the standard deviation of the sample mean and it is often called the *standard error* of the sample mean.

The quantity $\frac{S}{\sqrt{n}}$ is an estimate of the standard error of the sample mean, and is rightly called the *estimated standard error* of the sample mean although often the word *estimated* is left out.

T statistics occur often in many different problems when data is assumed normally distributed.

For example, in one and two sample t confidence intervals, regression and analysis of variance (to be studied in Modules 6 to 9) .

But they are *always* of the form of an *estimate of a population parameter* divided by its *(estimated) standard error.*

Example — t percentiles

Find the 97.5 percentiles of the t distribution with degrees of freedom 1,2,5, 30 and compare to that for the standard normal distribution.

```
qt(0.975, df = c(1, 2, 5, 30))

## [1] 12.706205  4.302653  2.570582  2.042272

qnorm(0.975)

## [1] 1.959964
```

Notes — t percentiles

The percentiles decrease as the degrees of freedom increase.

If the degrees of freedom are 30 or more the t percentile is increasingly close to the normal one,

Because the sample standard deviation is likely to be close to the population standard deviation,

And thus the difference between standardizing by the sample or population variance is small.

5 Central Limit Theorem & Law of Large Numbers — 5.6 , 5.7

5.1 Already Covered

Central Limit Theorem Already Covered

In Module 3 p. 32 and 33

But the main result is repeated here, with the variation that the sample variance replaces the population variance.

Previous slides in Module 5 covered samples with independent *normal* rv's.

Conclusions about a normally distributed sample mean still true with any distribution for *large* sample sizes and S replacing σ.

General guidance is that n should be about 30 to apply the central limit theorem but, especially if the underlying population has a symmetric distribution, sufficient accuracy can be obtained for smaller sample sizes.

Central Limit Theorem

Suppose $X_1, X_2, \cdots X_n$ are independent random variables all with the same distribution having mean μ and standard deviation σ.

Then letting $\bar{X}$ be the sample mean, $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$. From previous results,

$$E(\bar{X}) = \mu; \quad SD(\bar{X}) = \frac{\sigma}{\sqrt{n}}. \tag{9}$$

CLT: Central Limit Theorem says that for any number z,

$$P\left(\frac{\sqrt{n}(\bar{X}-\mu)}{S} \leq z\right) \text{ or } P\left(\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}) \leq z\right) \to P(Z \leq z) \tag{10}$$

as $n \to \infty$, where $Z \sim N(0,1)$.

Central Limit Theorem and Weak Law of Large Numbers

If the the number of observations size is large, then the distribution of the mean of independent random variables with the same distribution is approximately normal.

The approximate distribution is centred at the population mean, μ.

It has standard deviation $\propto$ both σ & S and inversely to $\sqrt{n}$.

The clustering of sample means around the population mean is the (Weak) Law of Large Numbers: no matter how small is a number $\epsilon > 0$,

$$P(|\bar{X} - \mu| > \epsilon) \to 0 \tag{11}$$

as $n \to \infty$.

The convergence in the Central Limit Theorem is called *Convergence in Distribution*, whereas the convergence in the Weak Law of Large Numbers is called *Convergence in Probability.*

Sample variance or population variance?

Because the sample size is large in the Opinion Poll example in Module 3 p.32 and 33, we could equally have used the sample variance — 0.36*0.64 — rather than the hypothesised population variance — 0.37*0.63.

5.2 Exponential rv's — the sample mean and rate

Example — Exponential

A good example of a skewed and very non-normal density is the exponential whose density $x > 0$ is $\lambda e^{-\lambda x}$. For this population, if a random sample of size n is taken:

1. What is the approximate probability $P(\sqrt{n}|\lambda\bar{X} - 1| \leq 2)$?
2. Plot the exact probability against n for $n = 1, 2, \ldots 30$.
3. Rewrite the event $[\sqrt{n}|\lambda\bar{X} - 1| > 2]$ so that λ is in the middle.

Solution — Exponential

The crucial point is that $\lambda X_1, \lambda X_2, \cdots \lambda X_n$ are exponential with parameter 1.
So the approximate probability is $P(|Z| \leq 2) = 0.9544997$.
The error between actual and approximate for $n = 30$ is 0.00166.

Solution — Exponential

The crucial point is that $\lambda X_1, \lambda X_2, \cdots \lambda X_n$ are exponential with parameter 1.
So the approximate probability is $P(|Z| \leq 2) = 0.9544997$.
The error between actual and approximate for $n = 30$ is 0.00166.
The next slide (Figure 2) shows the plot which follows from the R code:

```
n <- c(1:30)
actual_prob <- pgamma(n + 2 * sqrt(n), shape = n,
    rate = 1) - pgamma(n - 2 * sqrt(n), shape = n,
    rate = 1)
plot(n, actual_prob, type = "l")
abline(a = pnorm(2) - pnorm(-2), b = 0, lty = 2)
```

Sample size to 200 in Figure 3

Sample size to 500 in Figure 4

Solving for λ

Note that $|a| < b$ for $b > 0$ is the same as $-b < a < b$.

So $\sqrt{n}|\lambda\bar{X} - 1| < 2$ is the same as $-\frac{2}{\sqrt{n}} < \lambda\bar{X} - 1 < \frac{2}{\sqrt{n}}$,

And the latter is the same as $1 - \frac{2}{\sqrt{n}} < \lambda\bar{X} < 1 + \frac{2}{\sqrt{n}}$.

Dividing by $\bar{X}$ gives

$$[\sqrt{n}|\lambda\bar{X} - 1| < 2] = [\frac{1}{\bar{X}}(1 - \frac{2}{\sqrt{n}}) < \lambda < \frac{1}{\bar{X}}(1 + \frac{2}{\sqrt{n}})].$$

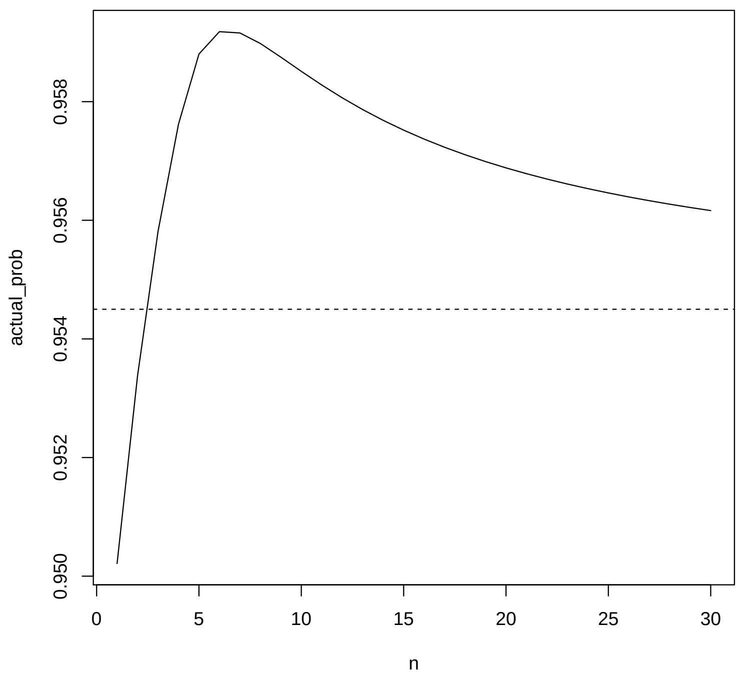

Figure 2: Actual Gamma probabilities with the normal approximation dotted

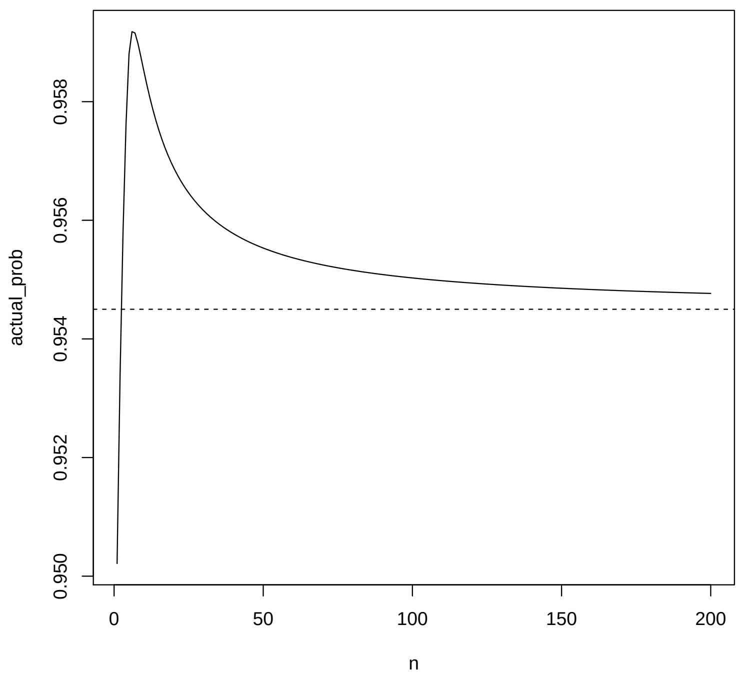

Figure 3: Actual Gamma probabilities with the normal approximation dotted

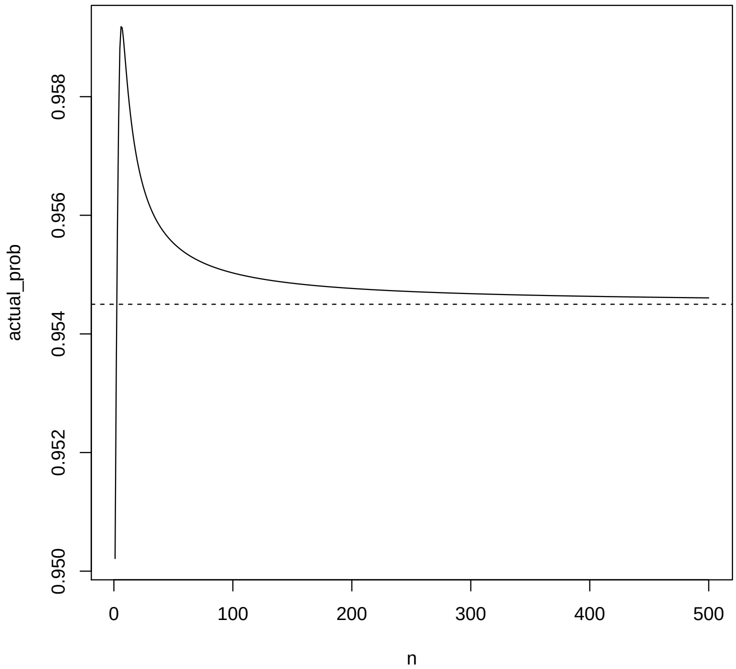

Figure 4: Actual Gamma probabilities with the normal approximation dotted

Observe that the event on the left has a probability that does not depend on λ, either exactly or approximately.

Also the event with λ in the middle is not an event about λ, because λ is a number not a random variable, but rather an event about the two random variables on the right and left of the inequality — this idea will be the centre of Module 7.