

Confidence intervals and the Feldman-Cousins construction

Edoardo Milotti

Advanced Statistics for Data Analysis

A.Y. 2015-16

Review of the Neyman construction of the confidence intervals

X—Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability

By J. NEYMAN

Reader in Statistics, University College, London

(Communicated by H. JEFFREYS, F.R.S.—Received 20 November, 1936—Read 17 June, 1937)

CONTENTS

	Page
I—INTRODUCTORY	333
(a) General Remarks, Notation, and Definitions	333
(b) Review of the Solutions of the Problem of Estimation Advanced Hereto	343
(c) Estimation by Unique Estimate and by Interval	346
II—CONFIDENCE INTERVALS	347
(a) Statement of the Problem	347
(b) Solution of the Problem of Confidence Intervals	350
(c) Example I	356
(d) Example II	362
(e) Family of Similar Regions Based on a Sufficient System of Statistics	364
(f) Example IIa	367
III—ACCURACY OF CONFIDENCE INTERVALS	370
(a) Shortest Systems of Confidence Intervals	370
(b) One-sided Estimation	374
(c) Example III	376
(d) Short Unbiased Systems of Confidence Intervals	377
IV—SUMMARY	378
V—REFERENCES	380

(b) *Review* of the Solutions of the Problem of Estimation Advanced Hereto*

The first attempt to solve the problem of estimation is connected with the theorem of Bayes and is applicable when the parameters $\theta_1, \theta_2, \dots, \theta_l$ in (5) are themselves random variables. The theorem of Bayes leads to the formula

$$p(\theta_1, \theta_2, \dots, \theta_l | x'_1, x'_2, \dots, x'_n) = \frac{p(\theta_1, \theta_2, \dots, \theta_l) p(x'_1, x'_2, \dots, x'_n | \theta_1, \dots, \theta_l)}{\int \dots \int p(\theta_1, \theta_2, \dots, \theta_l) p(x'_1, x'_2, \dots, x'_n | \theta_1, \dots, \theta_l) d\theta_1 \dots d\theta_l}, \quad (7).$$

representing the probability law of $\theta_1, \theta_2, \dots, \theta_l$, calculated under the assumption that the observations have provided the values (6) of the variables (4). Here $p(\theta_1, \dots, \theta_l)$ denotes the probability law of the θ 's, called *a priori*, and the integral in the denominator extends over all systems of values of the θ 's. The function $p(\theta_1, \theta_2, \dots, \theta_l | x'_1, x'_2, \dots, x'_n)$ is called the *a posteriori* probability law of θ 's. In cases where the *a priori* probability law $p(\theta_1, \theta_2, \dots, \theta_l)$ is known, the formula (7) permits the calculation of the most probable values of any of the θ 's and also of the probability that θ_i , say, will fall in any given interval, say, $a \leq \theta_i < b$. The most probable value of θ_i , say $\check{\theta}_i$, may be considered as the estimate of θ_i and then the probability, say

$$P\{\check{\theta}_i - \Delta < \theta_i < \check{\theta}_i + \Delta | E'\}, \quad \dots \dots \dots (8)$$

will describe the accuracy of the estimate $\check{\theta}_i$, where Δ is any fixed positive number and E' denotes the set (6) of observations.

It is known that, as far as we work with the conception of probability as adopted in this paper, the above theoretically perfect solution may be applied in practice only in quite exceptional cases, and this for two reasons :

(a) It is only very rarely that the parameters $\theta_1, \theta_2, \dots, \theta_l$ are random variables. They are generally unknown constants and therefore their probability law *a priori* has no meaning.

(b) Even if the parameters to be estimated, $\theta_1, \theta_2, \dots, \theta_l$, could be considered as random variables, the elementary probability law *a priori*, $p(\theta_1, \theta_2, \dots, \theta_l)$, is usually unknown, and hence the formula (7) cannot be used because of the lack of the necessary data.

II—CONFIDENCE INTERVALS

(a) *Statement of the Problem*

After these somewhat long preliminaries, we may proceed to the statement of the problem in its full generality.

Consider the variables (4) and assume that the form of their probability law (5) is known, that it involves the parameters $\theta_1, \theta_2, \dots, \theta_l$, which are constant (not random variables), and that the numerical values of these parameters are unknown. It is desired to estimate one of these parameters, say θ_1 . By this I shall mean that it is desired to define two functions $\underline{\theta}(E)$ and $\bar{\theta}(E) \supseteq \underline{\theta}(E)$, determined and single valued at any point E of the sample space, such that if E' is the sample point determined by observation, we can (1) calculate the corresponding values of $\underline{\theta}(E')$ and $\bar{\theta}(E')$, and (2) state that the true value of θ_1 , say θ_1^0 , is contained within the limits

$$\underline{\theta}(E') \leq \theta_1^0 \leq \bar{\theta}(E'), \dots \dots \dots (18)$$

this statement having some intelligible justification on the ground of the theory of probability.

This point requires to be made more precise. Following the routine of thought established under the influence of the Bayes Theorem, we could ask that, given the sample point E' , the probability of θ_1^0 falling within the limits (18) should be large, say, $\alpha = 0.99$, etc. If we express this condition by the formula

$$P\{\underline{\theta}(E') < \theta_1^0 < \bar{\theta}(E') | E'\} = \alpha, \dots \dots \dots (19)$$

we see at once that it contradicts the assumption that θ_1^0 is constant. In fact, on this assumption, whatever the fixed point E' and the values $\underline{\theta}(E')$ and $\bar{\theta}(E')$, the only values the probability (19) may possess are zero and unity. For this reason we shall drop the specification of the problem as given by the condition (19).

Returning to the inequalities (18), we notice that while the central part, θ_1^0 , is a constant, the extreme parts $\underline{\theta}(E')$ and $\bar{\theta}(E')$ are particular values of random variables. In fact, the coordinates of the sample point E are the random variables (4), and if $\underline{\theta}(E)$ and $\bar{\theta}(E)$ are single-valued functions of E , they must be random variables themselves.

Therefore, whenever the functions $\underline{\theta}(E)$ and $\bar{\theta}(E)$ are defined in one way or another, but the sample point E is not yet fixed by observation, we may legitimately discuss the probability of $\underline{\theta}(E)$ and $\bar{\theta}(E)$ fulfilling any given inequality and in particular the inequalities analogous to (18), in which, however, we must drop the dashes specifying a particular fixed sample point E' . We may also try to select $\underline{\theta}(E)$ and $\bar{\theta}(E)$ so that the probability of $\underline{\theta}(E)$ falling short of θ_1^0 and at the same time of $\bar{\theta}(E)$ exceeding θ_1^0 , is equal to any number α between zero and unity, fixed in advance. If θ_1^0 denotes the true value of θ_1 , then of course this probability must be calculated under the assumption that θ_1^0 is the true value of θ_1 . Thus we can look for two function $\underline{\theta}(E)$ and $\bar{\theta}(E)$, such that

$$P \{ \underline{\theta}(E) \leq \theta_1^0 \leq \bar{\theta}(E) | \theta_1^0 \} = \alpha. \quad (20)$$

and require that the equation (20) holds good *whatever* the value θ_1^0 of θ_1 and *whatever* the values of the other parameters $\theta_2, \theta_3, \dots, \theta_l$, involved in the probability law of the X 's may be.

The functions $\underline{\theta}(E)$ and $\bar{\theta}(E)$ satisfying the above conditions will be called the lower and the upper confidence limits of θ_1 . The value α of the probability (20) will be called the confidence coefficient, and the interval, say $\delta(E)$, from $\underline{\theta}(E)$ to $\bar{\theta}(E)$, the confidence interval corresponding to the confidence coefficient α .

It is obvious that the form of the functions $\underline{\theta}(E)$ and $\bar{\theta}(E)$ must depend upon the probability law $p(E | \theta_1 \dots \theta_l)$.

(b) *Solution of the Problem of Confidence Intervals*

In order to find the solution of the problem of confidence intervals, let us suppose that it is already solved and that $\underline{\theta} (E)$ and $\bar{\theta} (E)$ are functions determined and single valued in the whole sample space, W , and such that the equality (20) holds good whatever the true values of the parameters $\theta_1, \theta_2, \dots \theta_l$. It will be convenient to interpret the situation geometrically. For this purpose we shall need to consider the space, G , of $n + 1$ dimensions which we shall call the general space. The points in this space will be determined by $n + 1$ coordinates $x_1, x_2, \dots x_n, \theta_1$, the first n of which are the particular values of the random variables (4) and thus determine the position of a sample point, E , in the n -dimensional space W , and the last coordinate θ_1 is one of the possible values of the parameter θ_1 in the probability law $p (E|\theta_1 \dots \theta_l)$ which we desire to estimate.

Consequently, if we consider any hyperplane, $G (\theta_1)$ in G corresponding to the equation $\theta_1 = \text{const.}$, this may be interpreted as an image of the sample space W . We notice also that to any point E in the sample space W there will correspond in G a straight line, say $L (E)$, parallel to the axis $O\theta_1$. If $x_1', x_2' \dots x_n'$ are the coordinates of E' , then the line $L (E')$ will correspond to the equations $x_i = x_i'$ for $i = 1, 2, \dots n$.

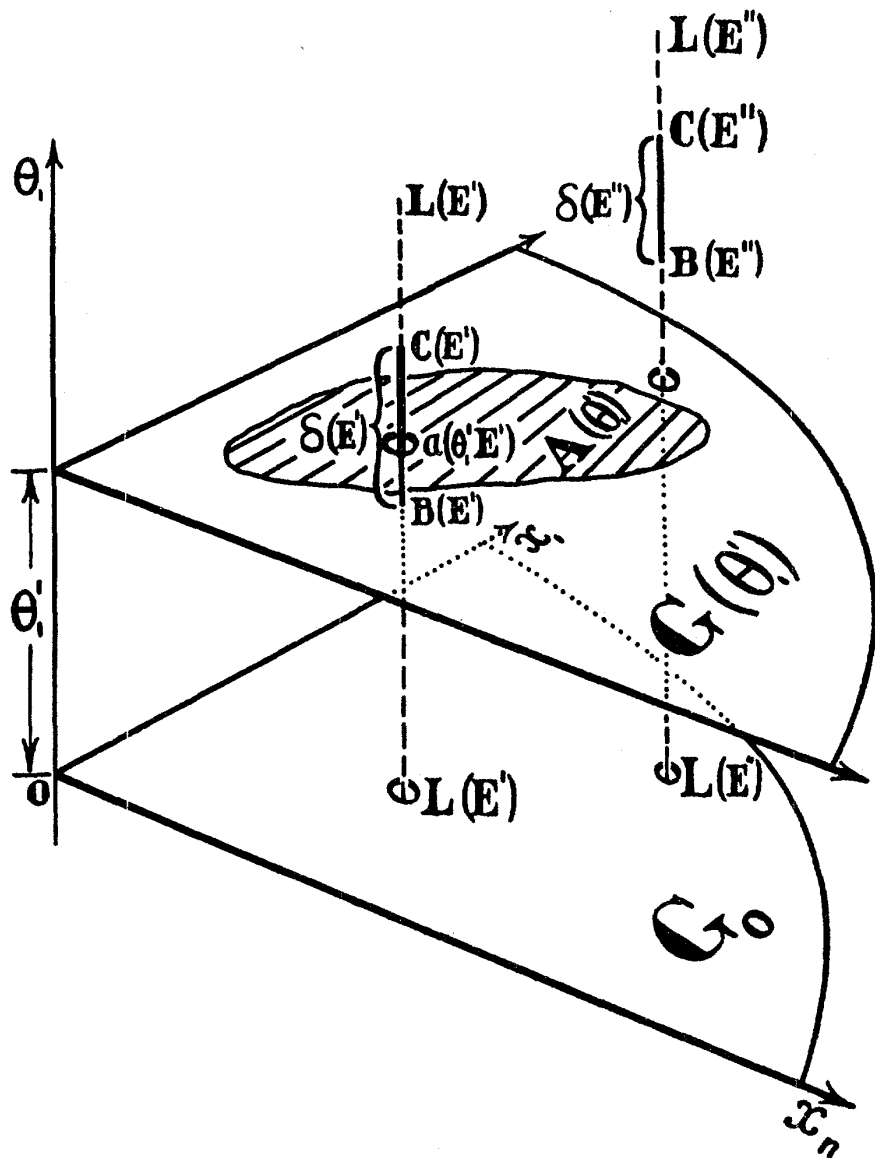
Consider now the functions $\underline{\theta} (E)$ and $\bar{\theta} (E)$. On each line $L (E)$, they will determine two points, say $B (E)$ and $C (E)$ with coordinates

$$x_1, x_2, \dots x_n, \underline{\theta} (E) \dots \dots \dots (22)$$

and

$$x_1, x_2, \dots x_n, \bar{\theta} (E) \dots \dots \dots (23)$$

respectively, where $x_1, x_2, \dots x_n$ are the coordinates of the sample point E . The interval between $B (E)$ and $C (E)$ will be the image of the confidence interval $\delta (E)$ corresponding to the sample point E . If we fix a value of $\theta_1 = \theta_1'$ and a sample point E' , then the hyperplane $G (\theta_1')$ may cut or may not cut the confidence interval $\delta (E')$. If $G (\theta_1')$ does cut $\delta (E')$, let $a (\theta_1', E')$ denote the point of intersection.



- vertical axis: parameter
- other axes: data

The acceptance region A represents the set of covering confidence intervals for each particular dataset.

Each confidence interval is defined by

$$\underline{\theta}(E) \leq \theta'_1 \leq \bar{\theta}(E). \quad \dots \quad (24)$$

and corresponds to a given confidence level (a given probability of containing the true value of the parameter)

FIG. 1—The general space G.

The conception and properties of the regions of acceptance are exceedingly important from the point of view of the theory given below. We shall therefore discuss them in detail proving separately a few propositions, however simple they may seem to be.

Proposition I—Whenever the sample point E falls within the region of acceptance $A(\theta'_1)$, corresponding to any fixed value θ'_1 of θ_1 , then the corresponding confidence interval $\delta(E)$ must cover θ'_1 .

Proof—This proposition is a direct consequence of the definition of the region of acceptance. Suppose it is not true. Then there must be at least one sample point, say E' , which falls within $A(\theta'_1)$ and such that either $\underline{\theta}(E') \leq \bar{\theta}(E') < \theta'_1$ or $\theta'_1 < \underline{\theta}(E') \leq \bar{\theta}(E')$. Comparing these inequalities with (24) which serve as a definition of the region of acceptance $A(\theta'_1)$, we see that E' could not fall within $A(\theta'_1)$, which proves the Proposition I.

Proposition II—If a confidence interval $\delta (E'')$ corresponding to a sample point E'' covers a value θ'_1 of θ_1 , then the sample point E'' must fall within $A (\theta'_1)$.

Proof—If $\delta (E'')$ covers θ'_1 , then it follows that $\underline{\theta} (E'') \leq \theta'_1 \leq \bar{\theta} (E'')$. Comparing these inequalities with (24) defining the region $A (\theta'_1)$, we see that E'' must fall within $A (\theta'_1)$.

If we agree to denote generally by $\{B \in A\}$ the words “ B belongs to A ” or “ B is an element of A ”, then we may sum up the above two propositions by writing the identity

$$\{E \in A (\theta'_1)\} \equiv \{\delta (E) \subset \theta'_1\} \equiv \{\underline{\theta} (E) \leq \theta'_1 \leq \bar{\theta} (E)\}, \quad . . . \quad (25)$$

meaning that the event consisting in the sample point E falling within the region of acceptance $A (\theta'_1)$ is equivalent to the other event which consists in θ'_1 being covered by $\delta (E)$.

Corollary I—It follows from the Proposition I and II that whatever may be the true values $\theta'_1, \theta'_2 \dots \theta'_l$ of the θ 's, the probability of any fixed value θ''_1 of θ_1 being covered by $\delta (E)$ is identical with the probability of the sample point E falling within $A (\theta''_1)$.

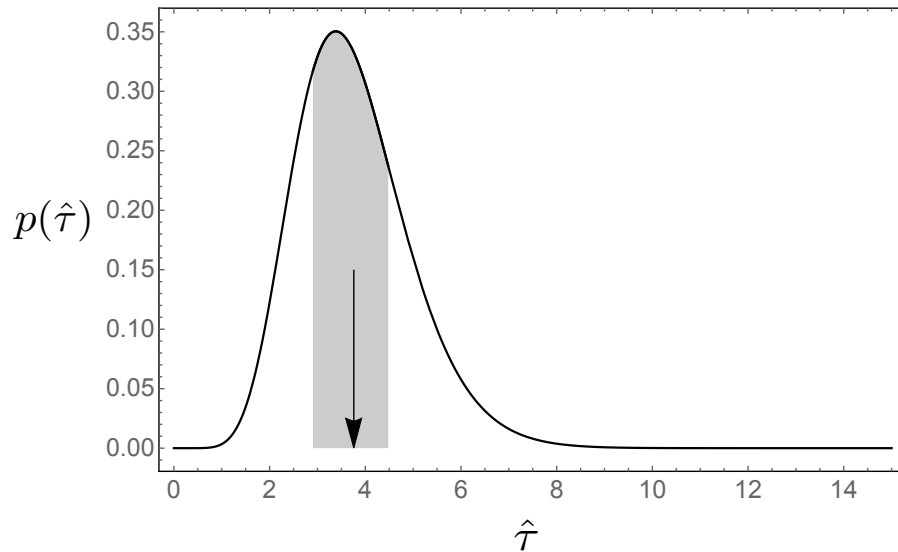
$$\begin{aligned} P \{\delta (E) \subset \theta''_1 | \theta'_1, \dots \theta'_l\} &= P \{\underline{\theta} (E) \leq \theta''_1 \leq \bar{\theta} (E) | \theta'_1, \theta'_2, \dots \theta'_l\} \\ &= P \{E \in A (\theta''_1) | \theta'_1, \theta'_2, \dots \theta'_l\}. \end{aligned} \quad (26)$$

Review of the Neyman construction of the confidence intervals

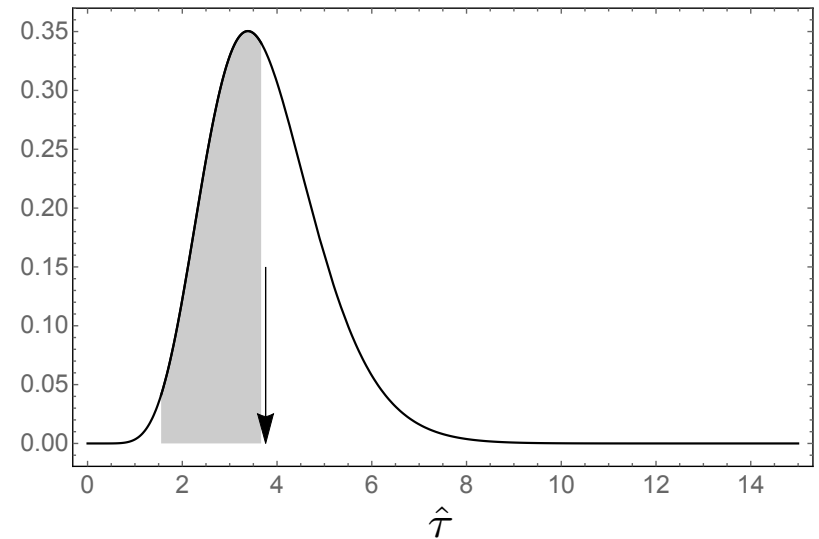
1. Arbitrariness of confidence intervals

Example: estimate of the mean decay time in an exponential distribution with

$$\tau = 3.76; \quad n = 10$$



Covering (central) interval

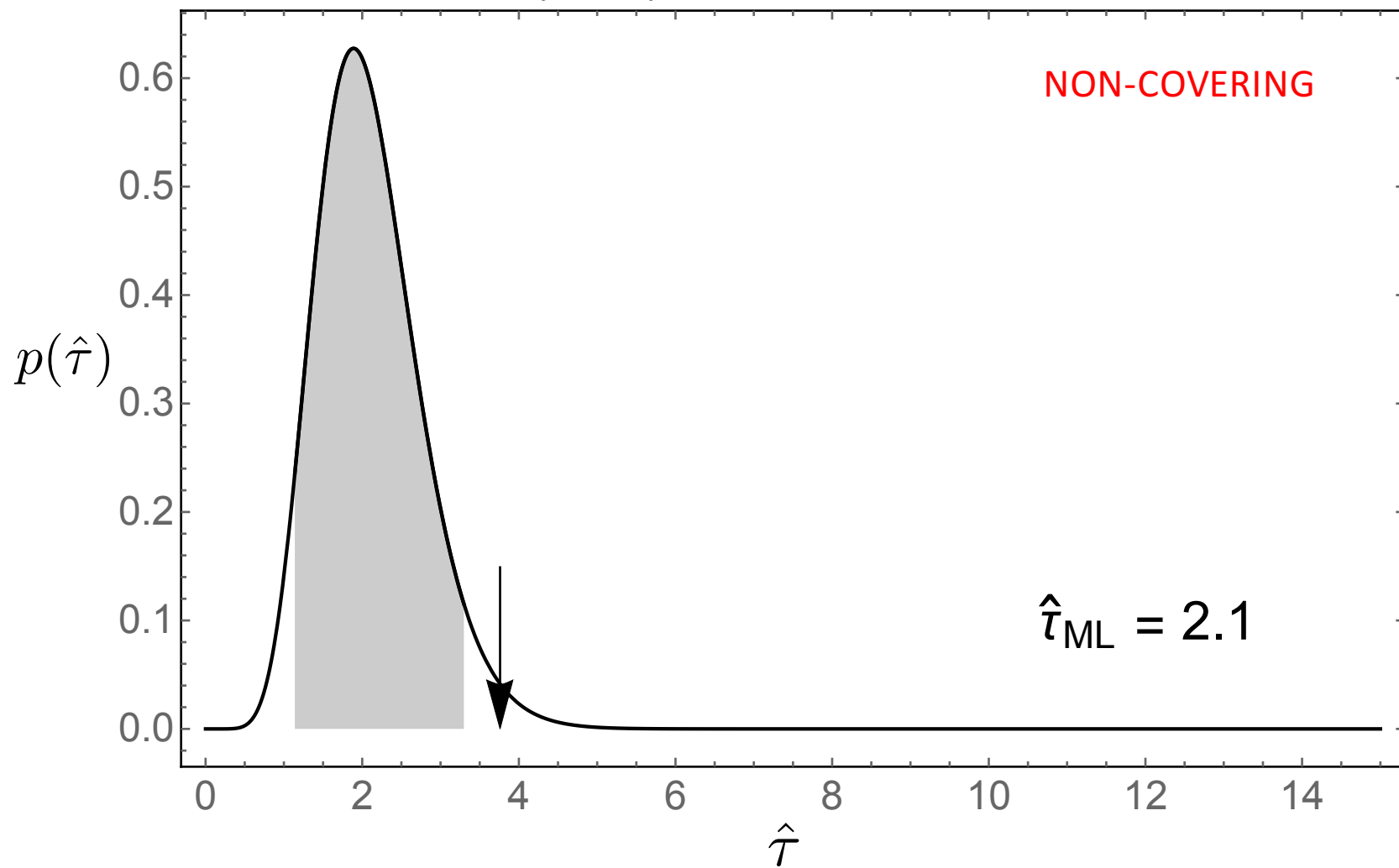


Non-covering interval

In these examples the confidence level (probability corresponding to the shaded area) = 0.5

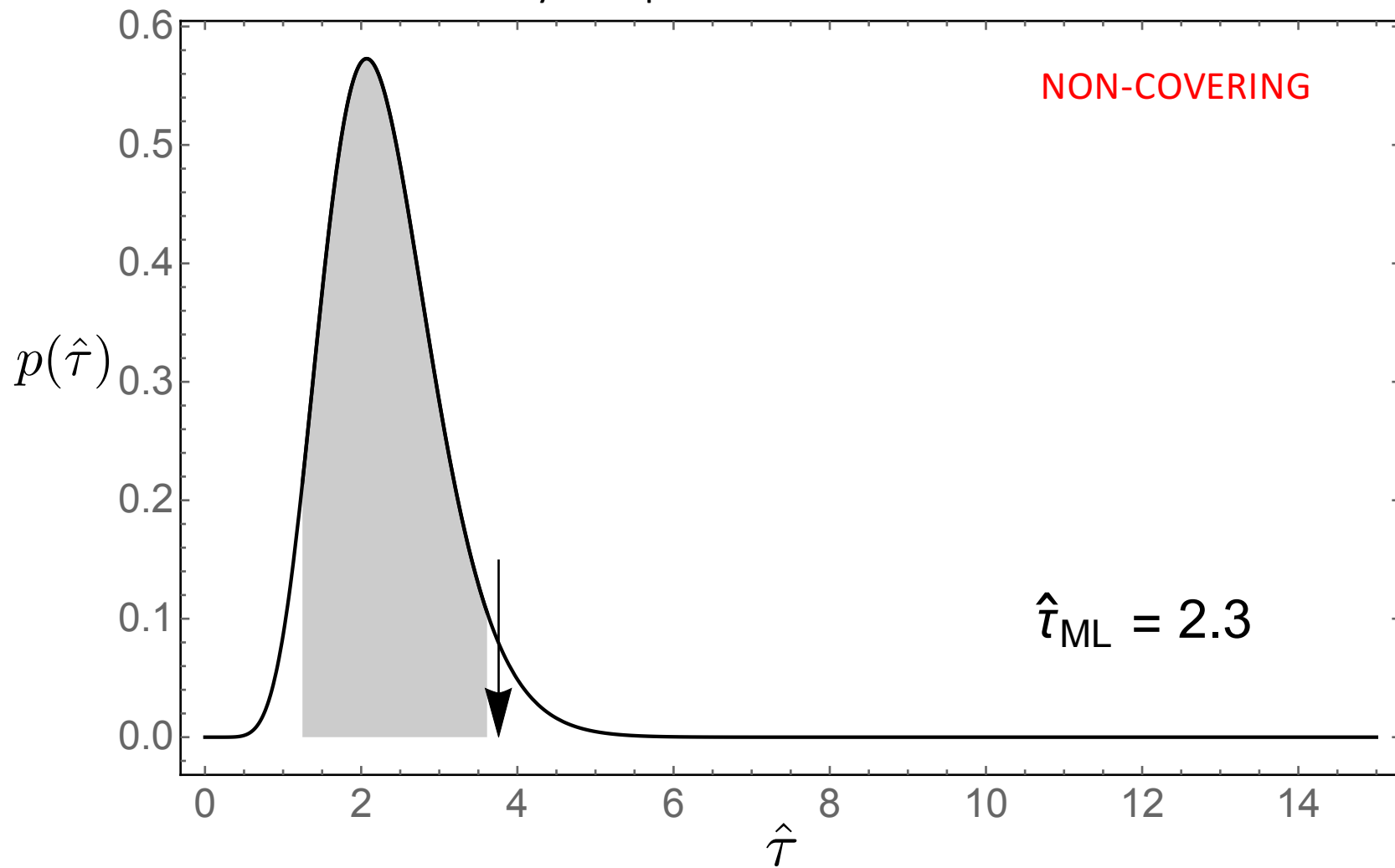
2. Central intervals as random variables

Different estimates of the decay time produce different central intervals



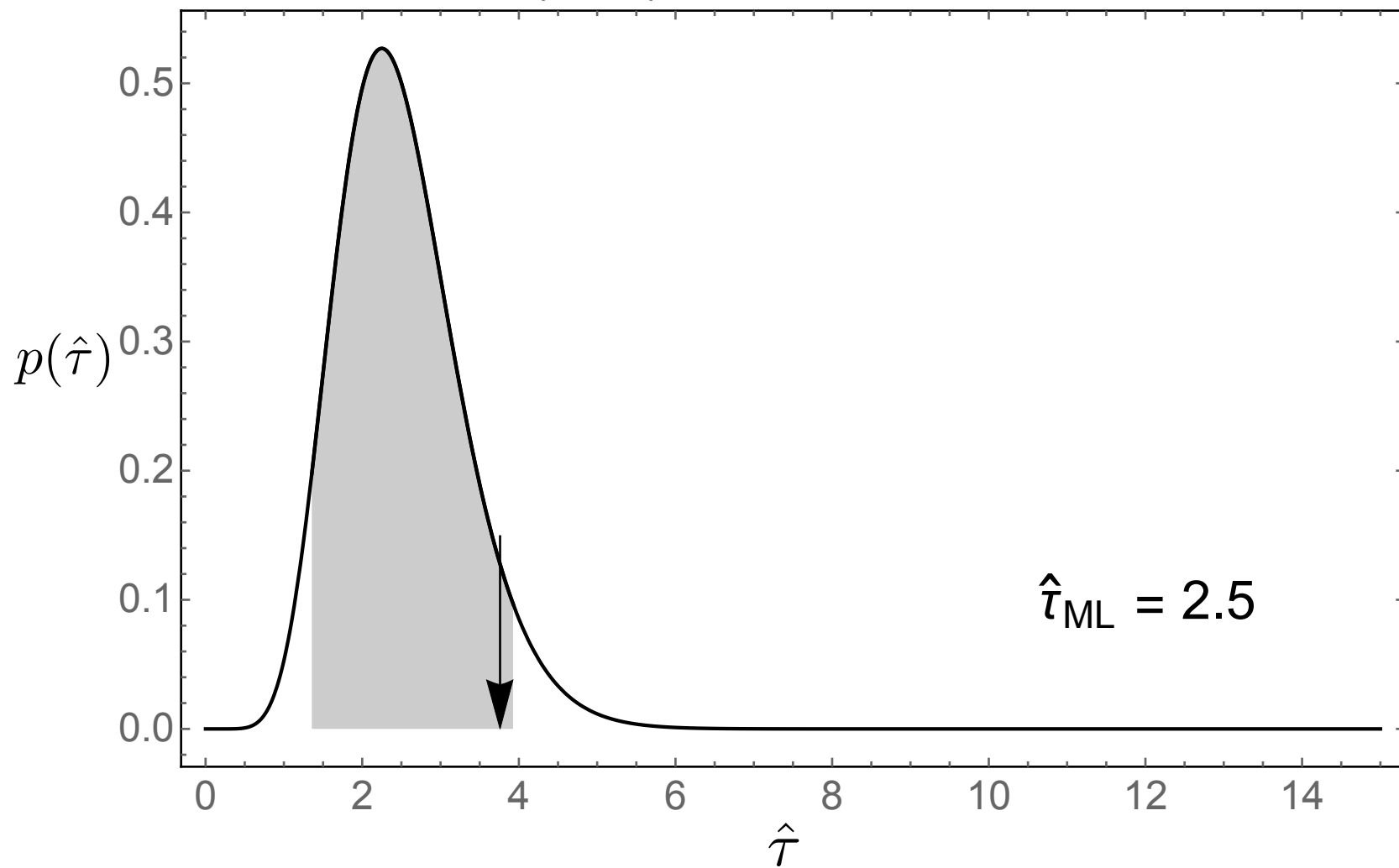
2. Central intervals as random variables

Different estimates of the decay time produce different central intervals



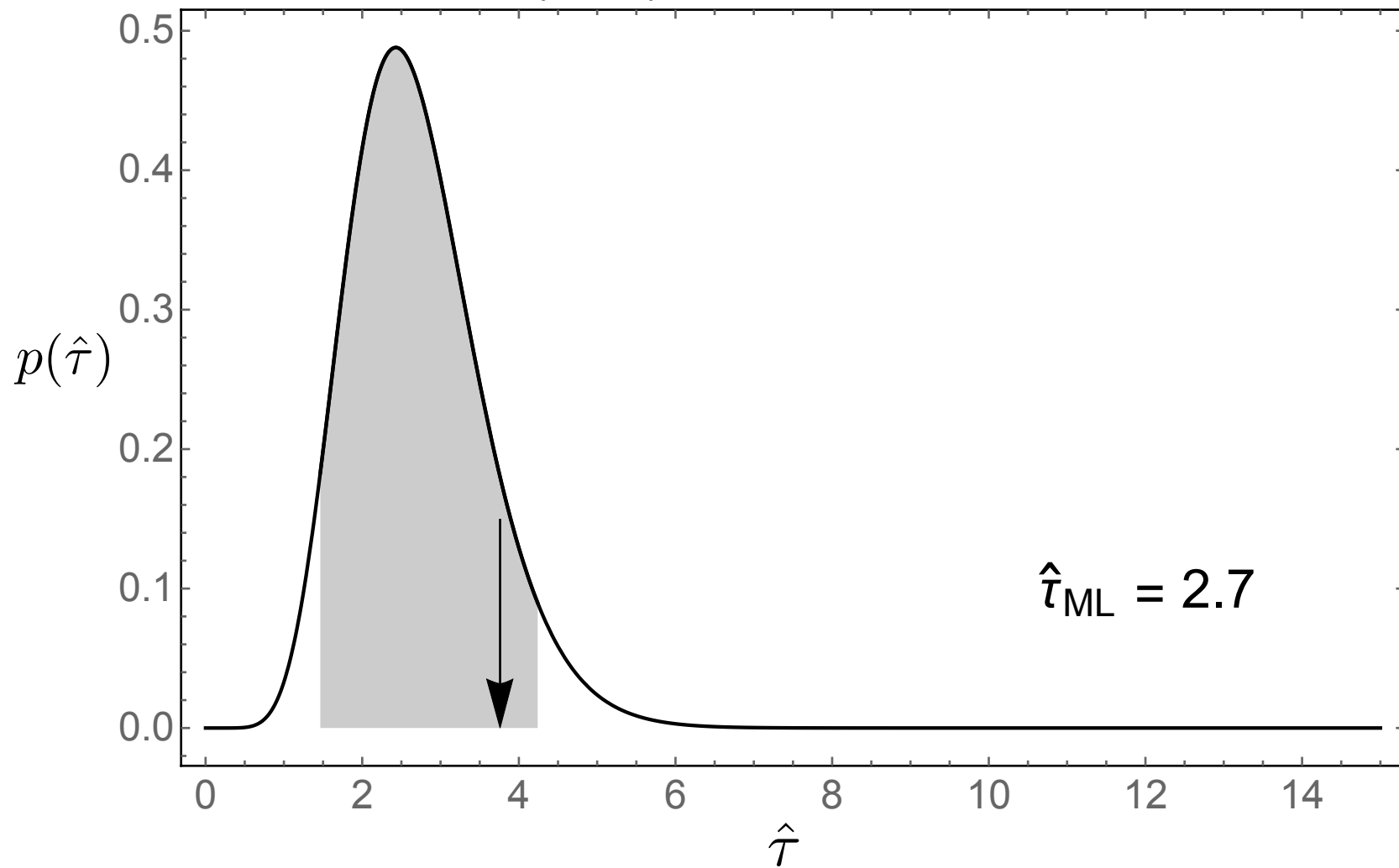
2. Central intervals as random variables

Different estimates of the decay time produce different central intervals



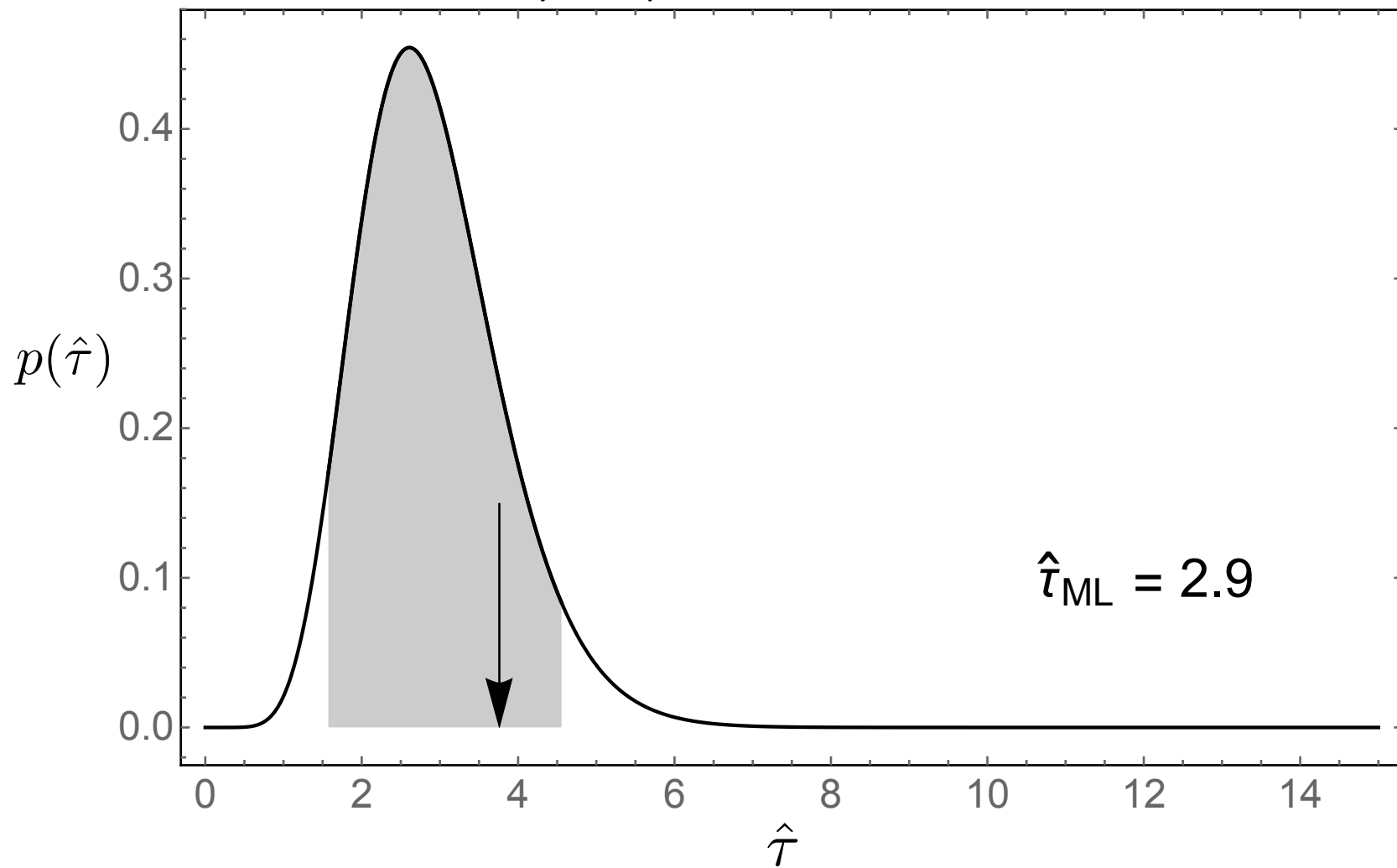
2. Central intervals as random variables

Different estimates of the decay time produce different central intervals



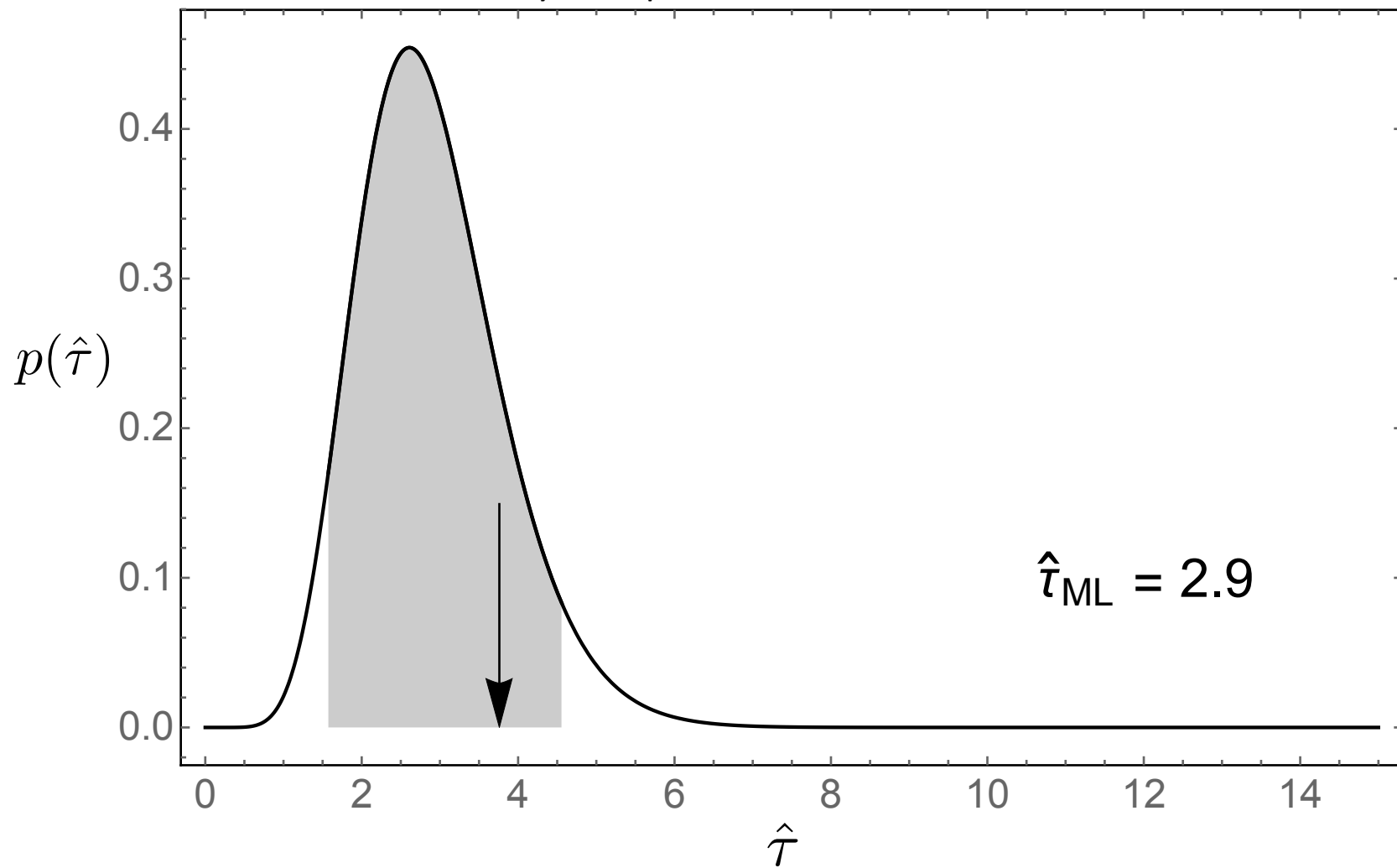
2. Central intervals as random variables

Different estimates of the decay time produce different central intervals



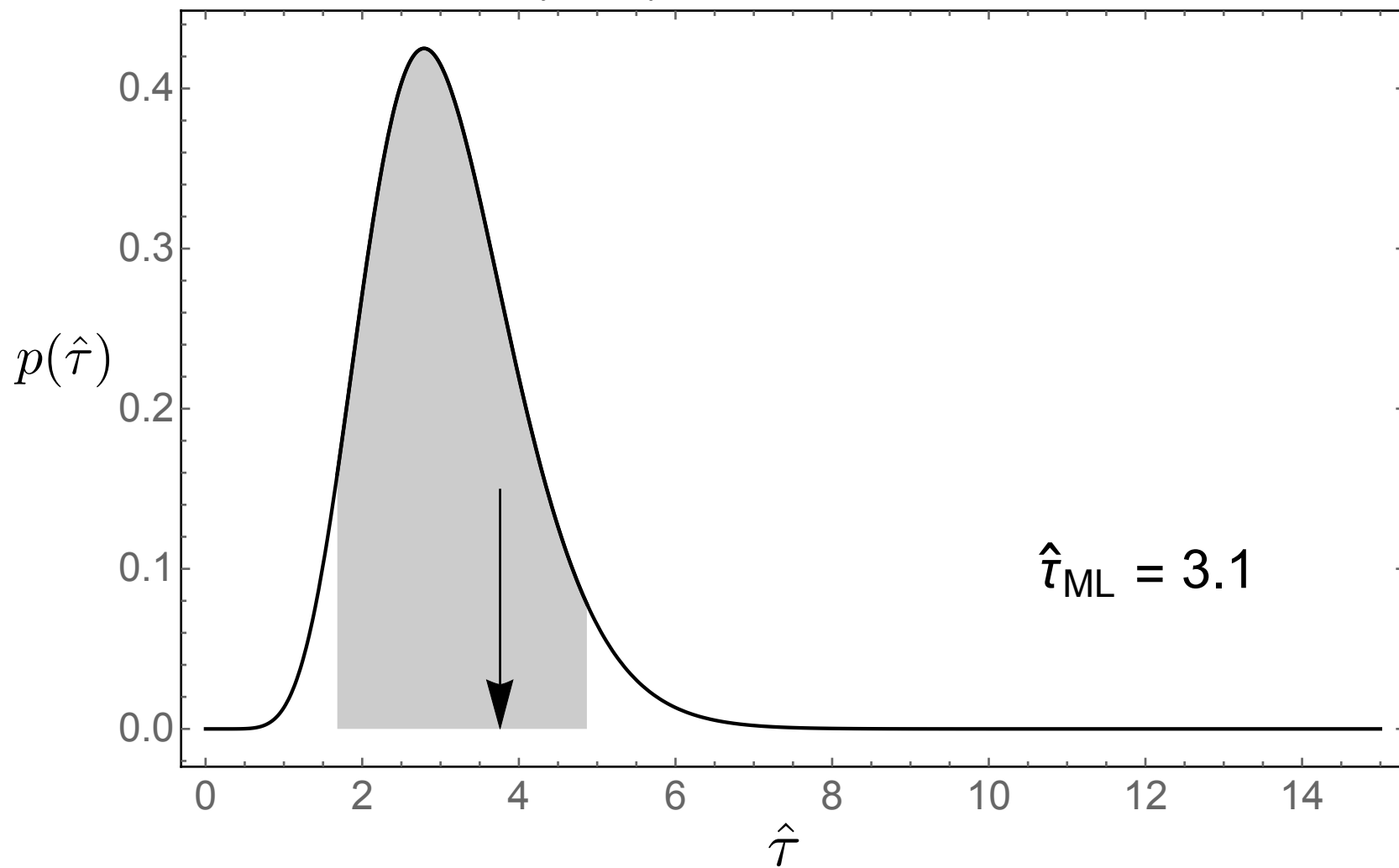
2. Central intervals as random variables

Different estimates of the decay time produce different central intervals



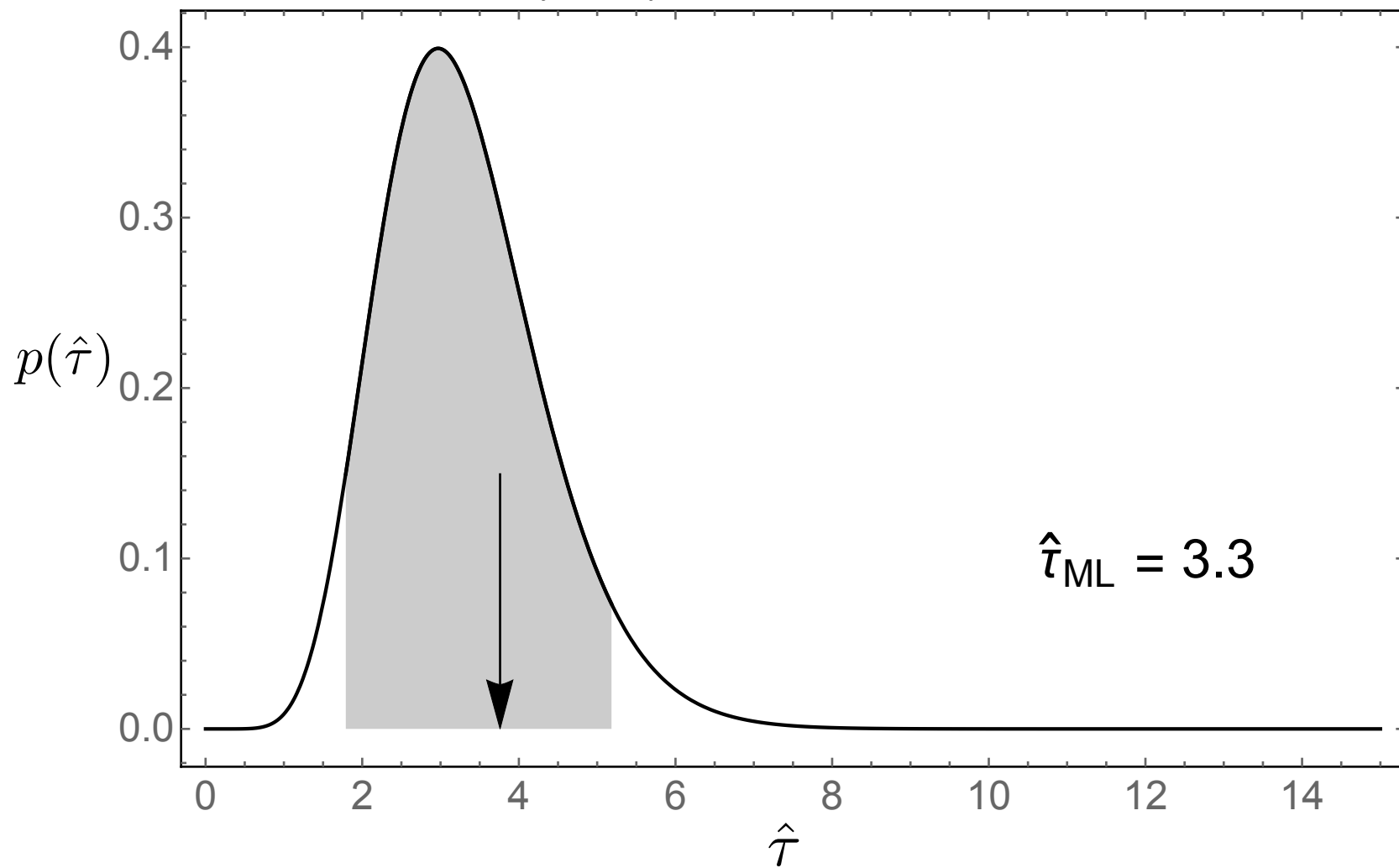
2. Central intervals as random variables

Different estimates of the decay time produce different central intervals



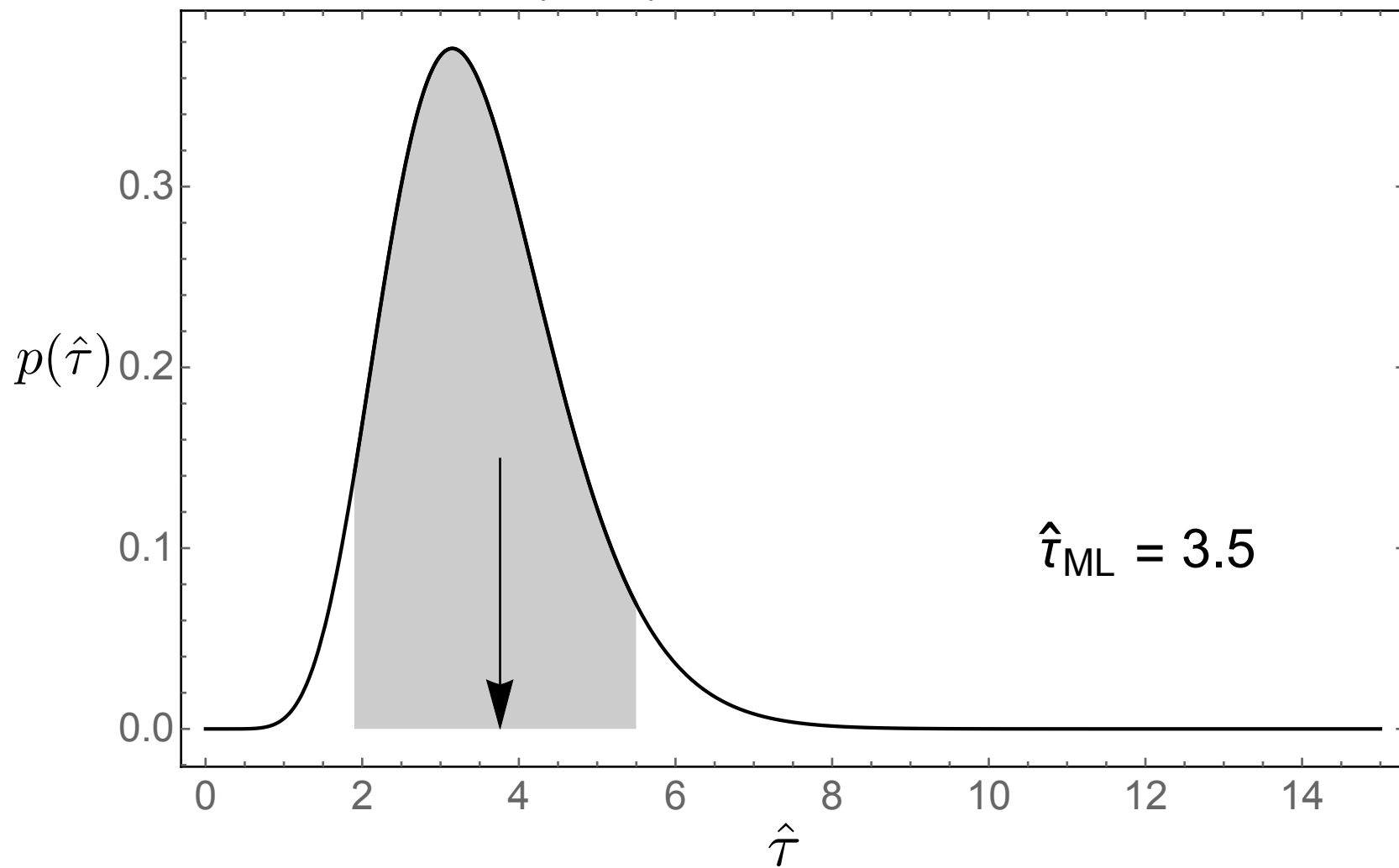
2. Central intervals as random variables

Different estimates of the decay time produce different central intervals



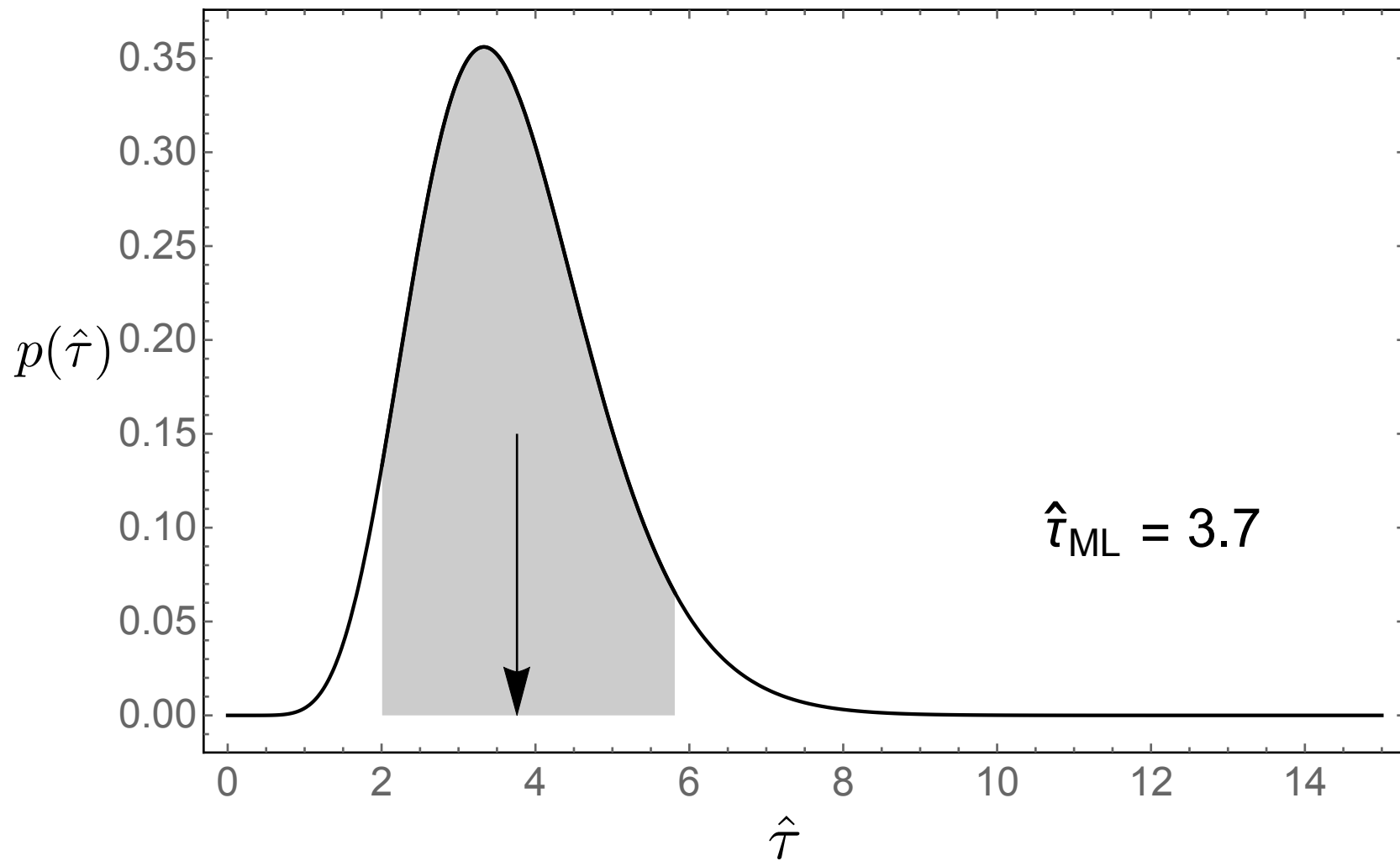
2. Central intervals as random variables

Different estimates of the decay time produce different central intervals



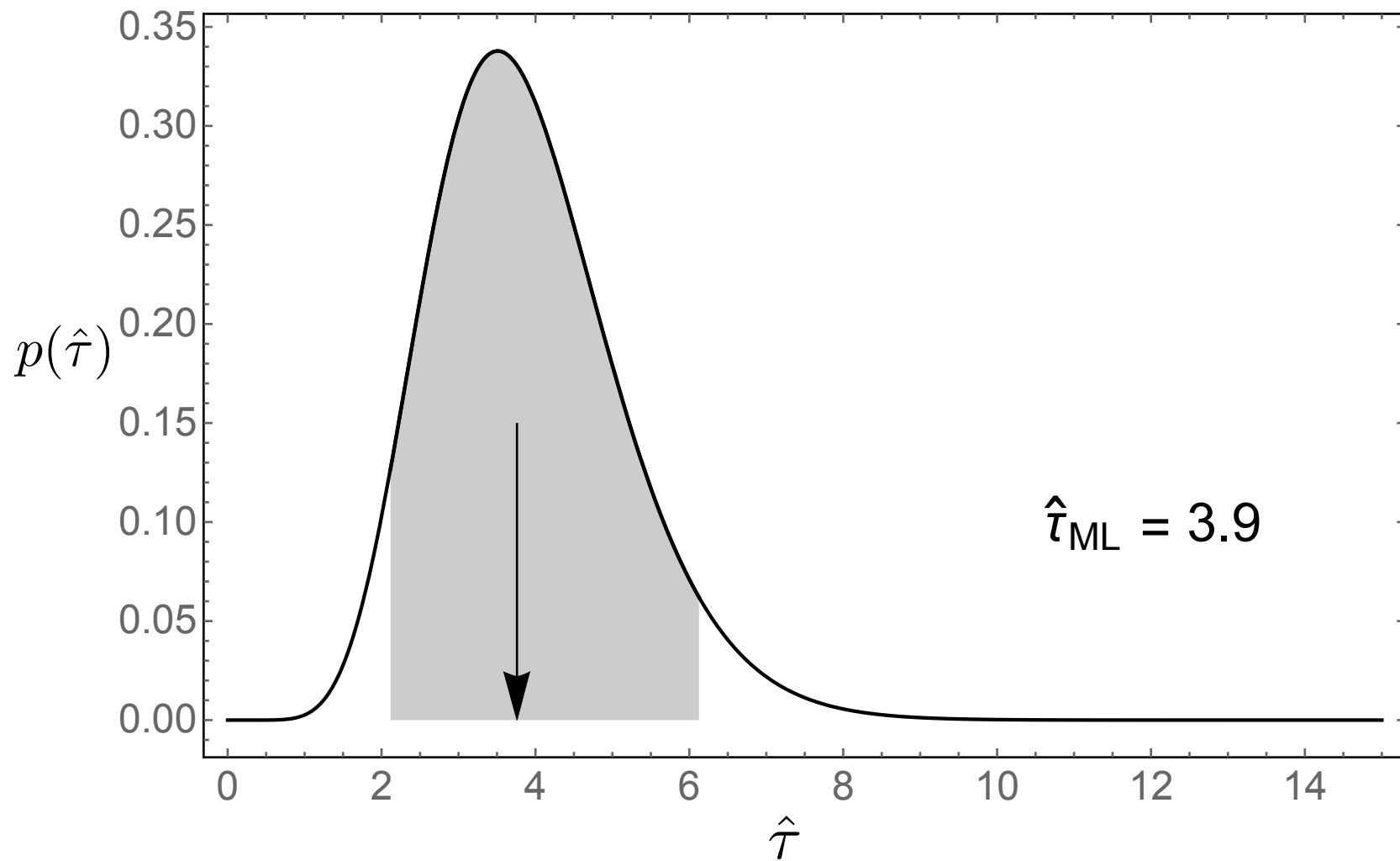
2. Central intervals as random variables

Different estimates of the decay time produce different central intervals



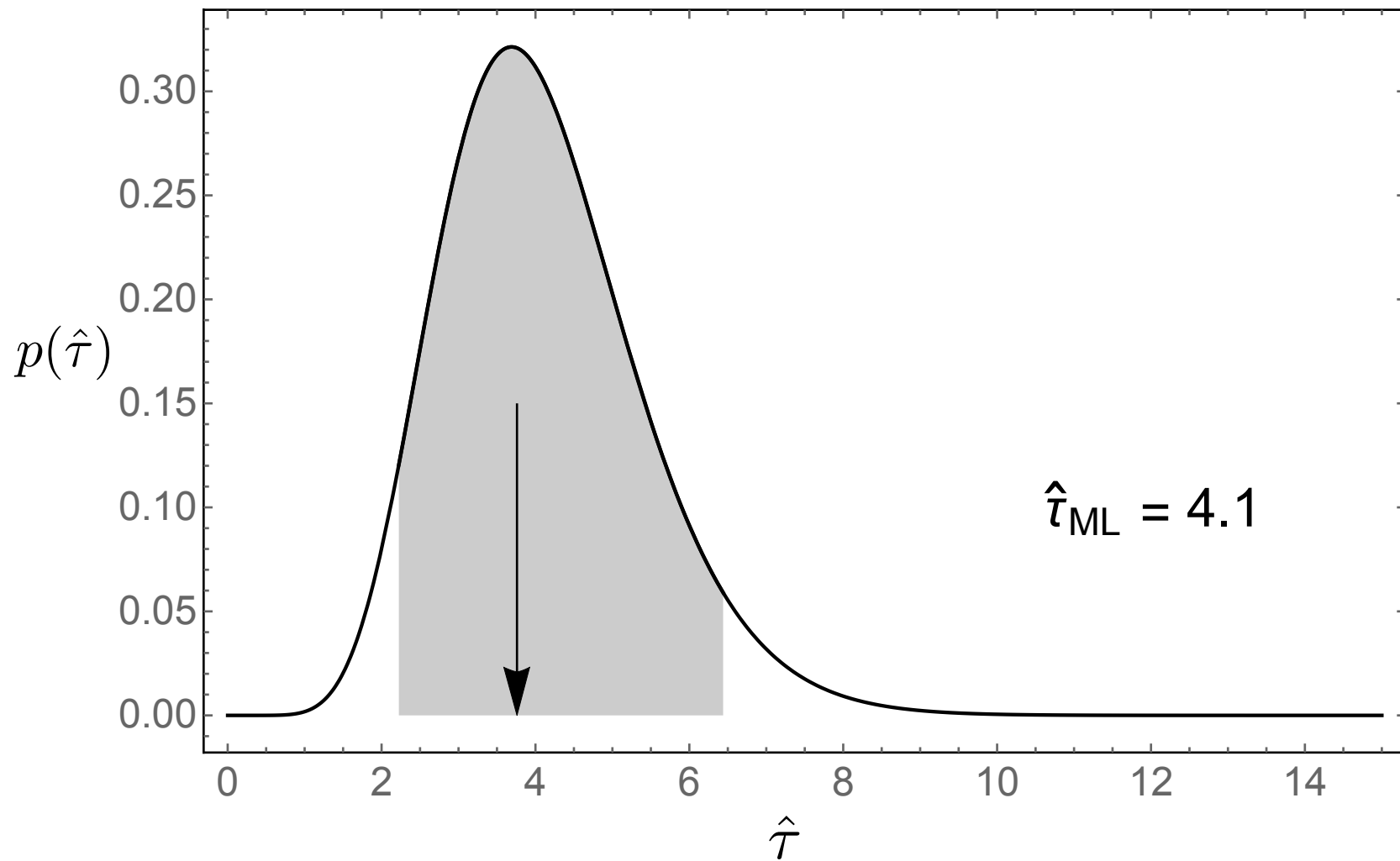
2. Central intervals as random variables

Different estimates of the decay time produce different central intervals



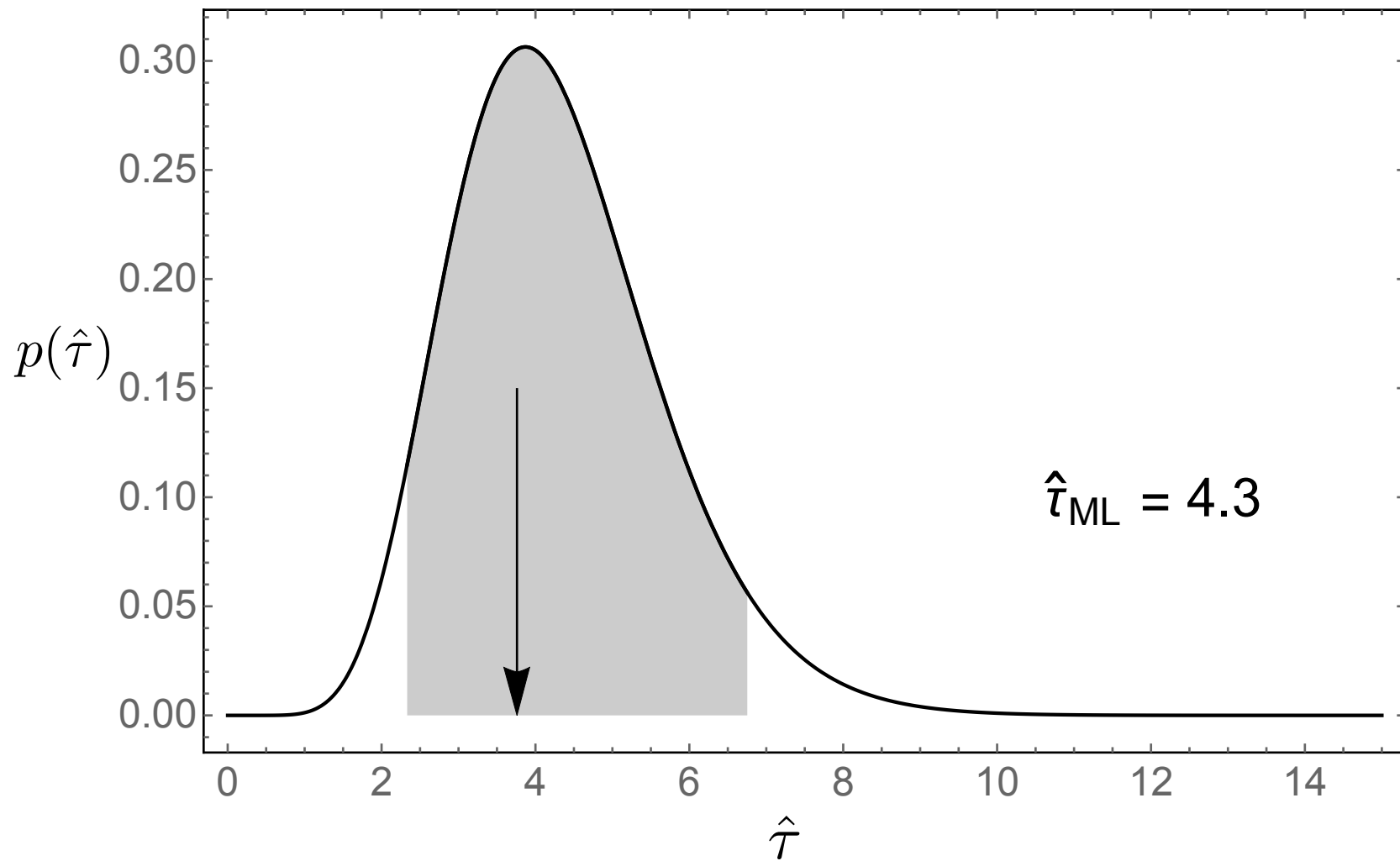
2. Central intervals as random variables

Different estimates of the decay time produce different central intervals



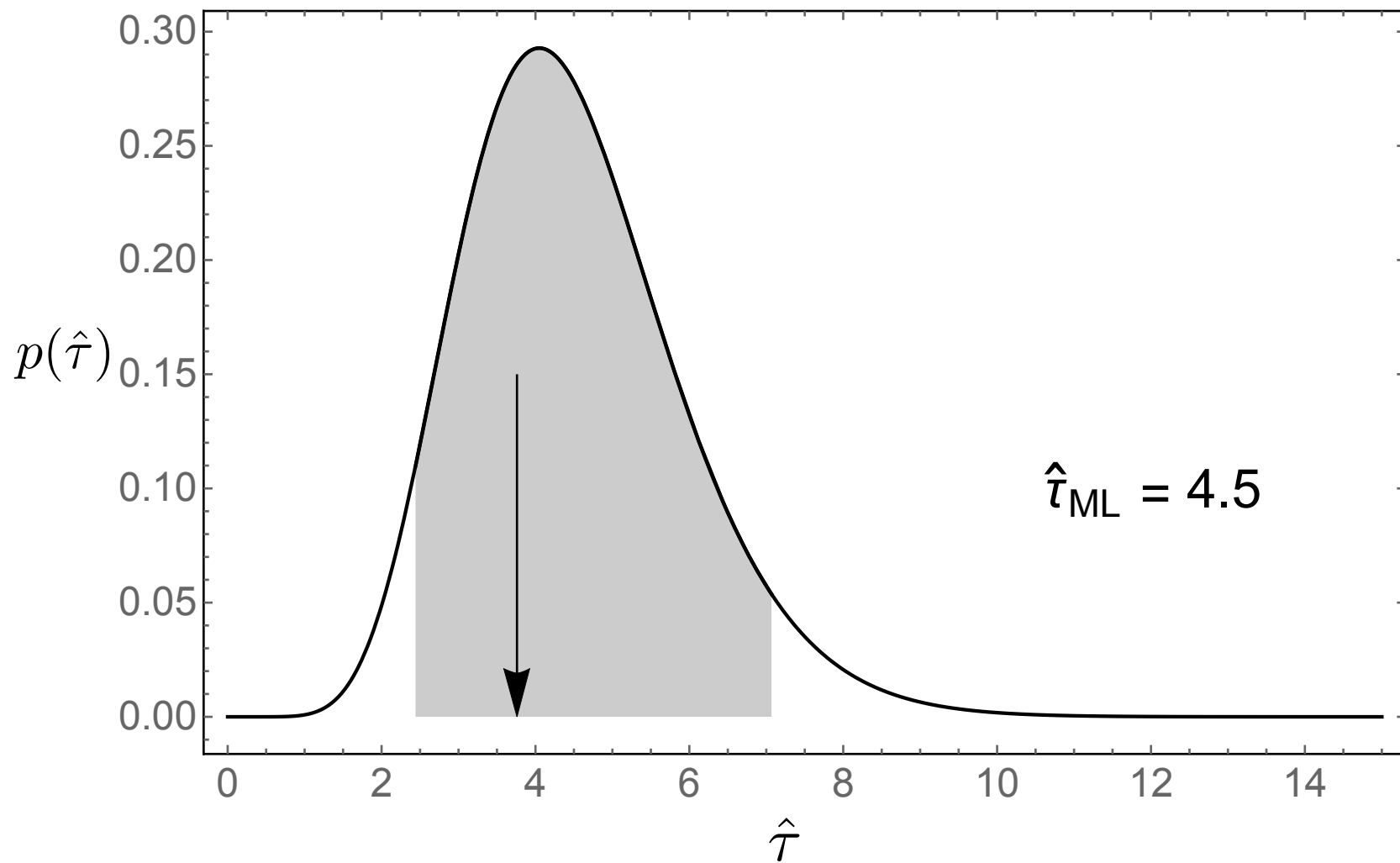
2. Central intervals as random variables

Different estimates of the decay time produce different central intervals



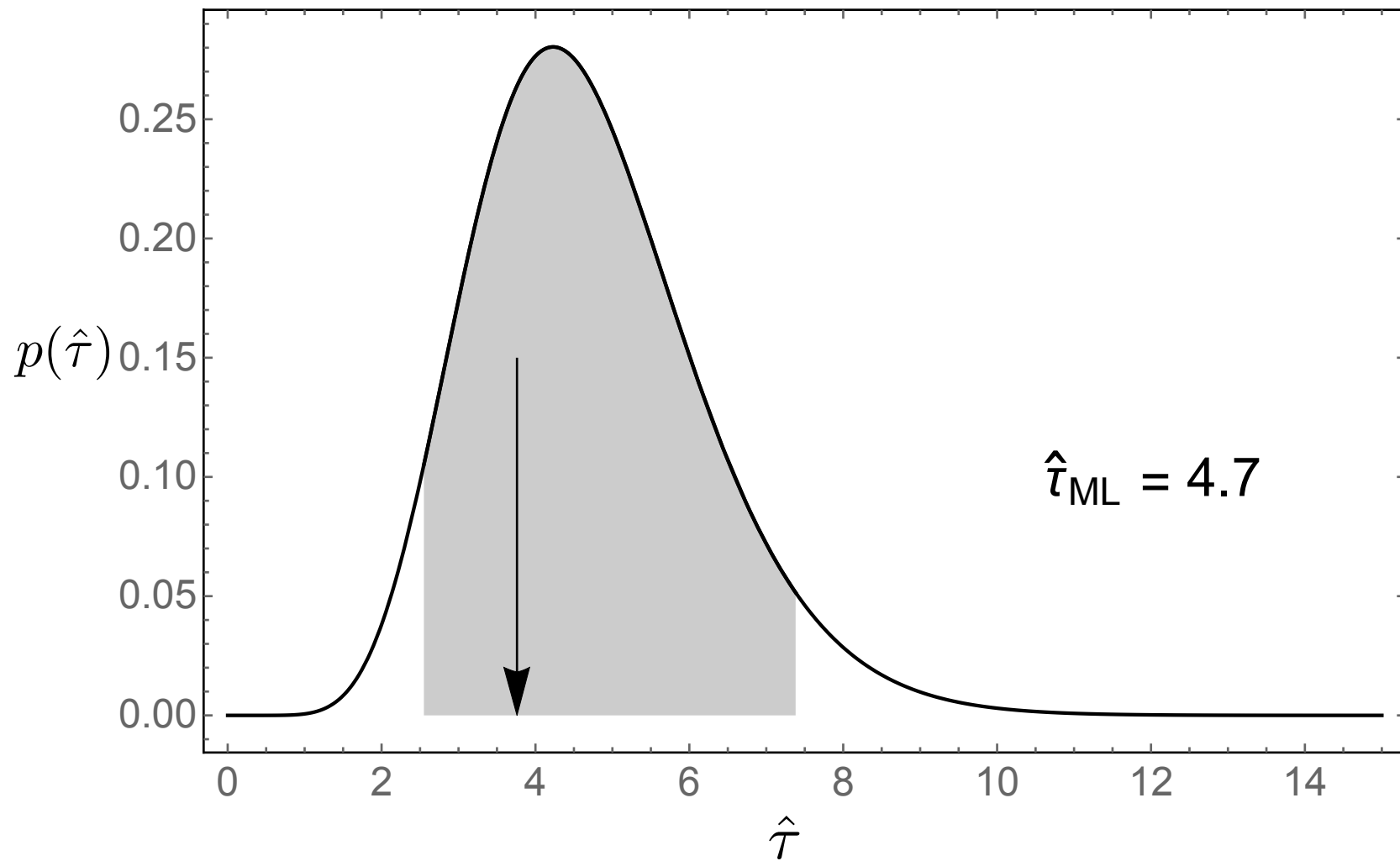
2. Central intervals as random variables

Different estimates of the decay time produce different central intervals



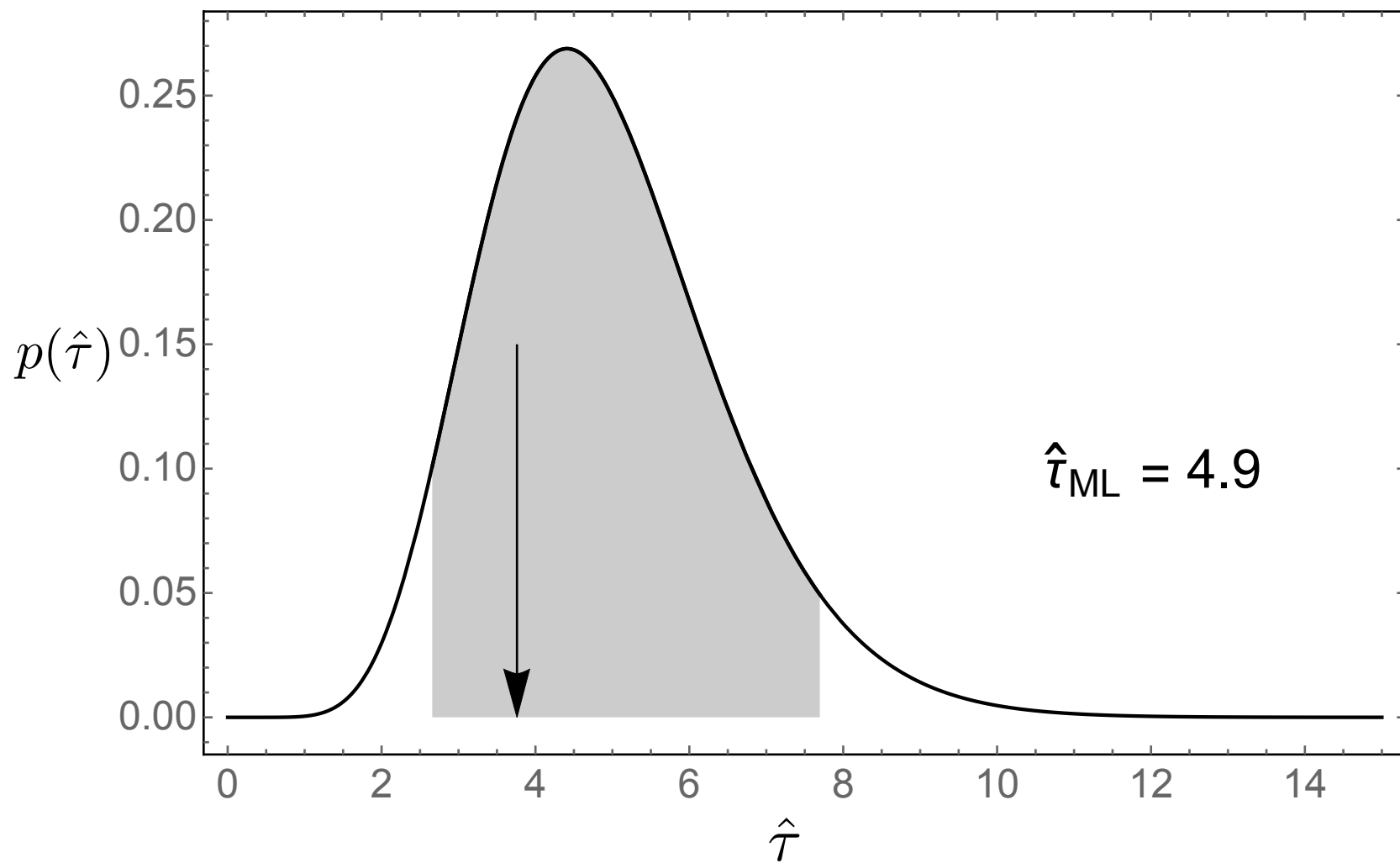
2. Central intervals as random variables

Different estimates of the decay time produce different central intervals



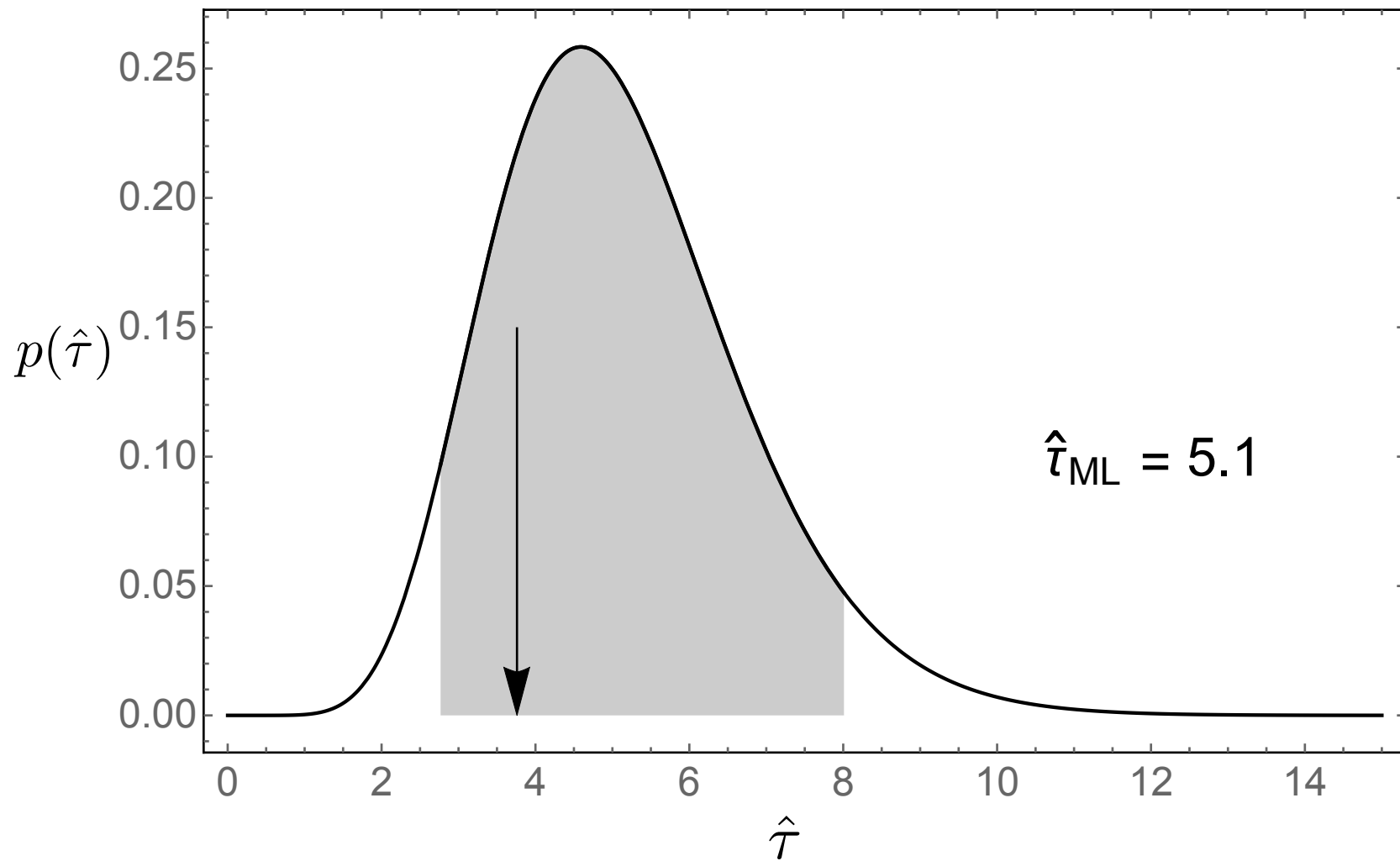
2. Central intervals as random variables

Different estimates of the decay time produce different central intervals



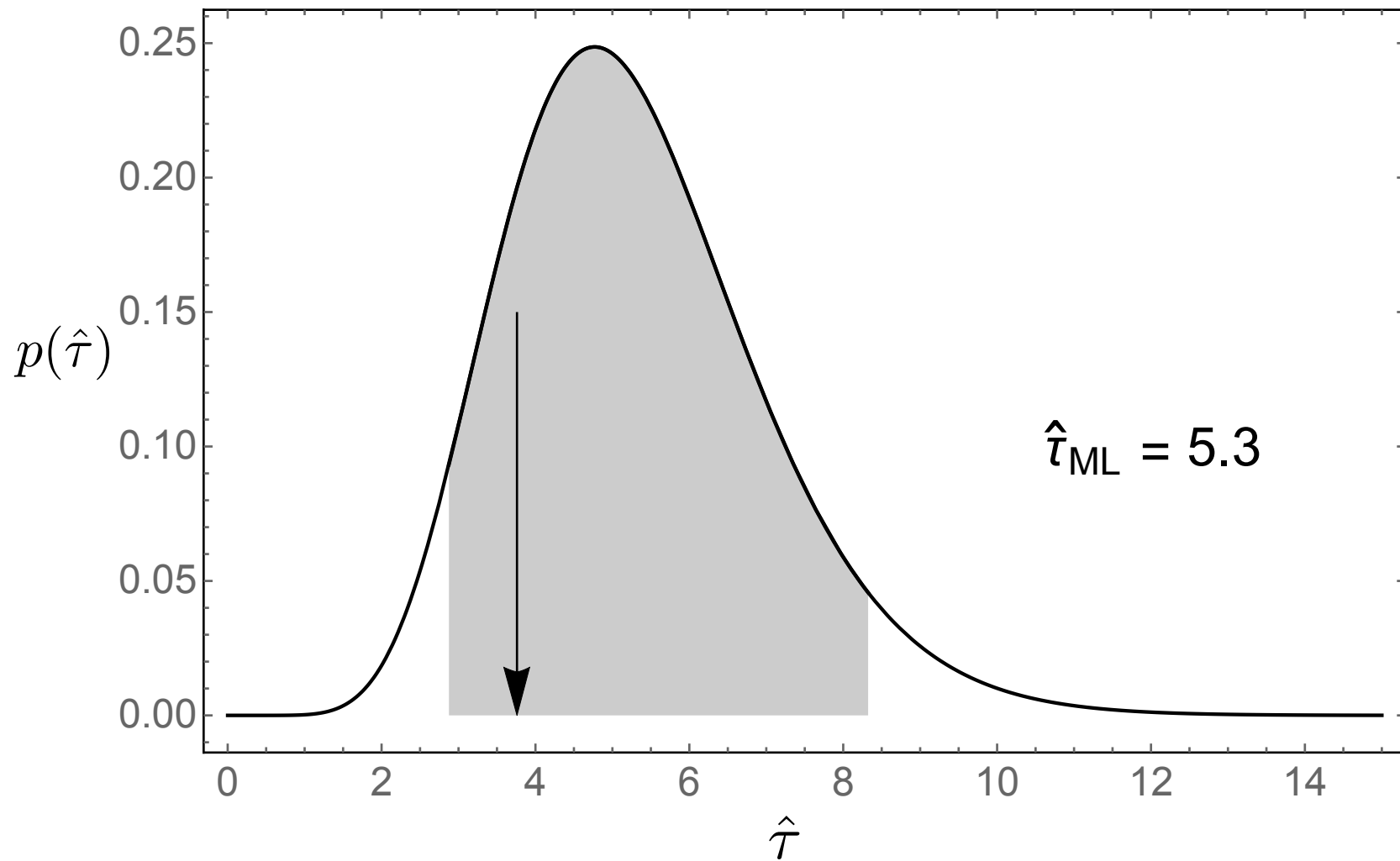
2. Central intervals as random variables

Different estimates of the decay time produce different central intervals



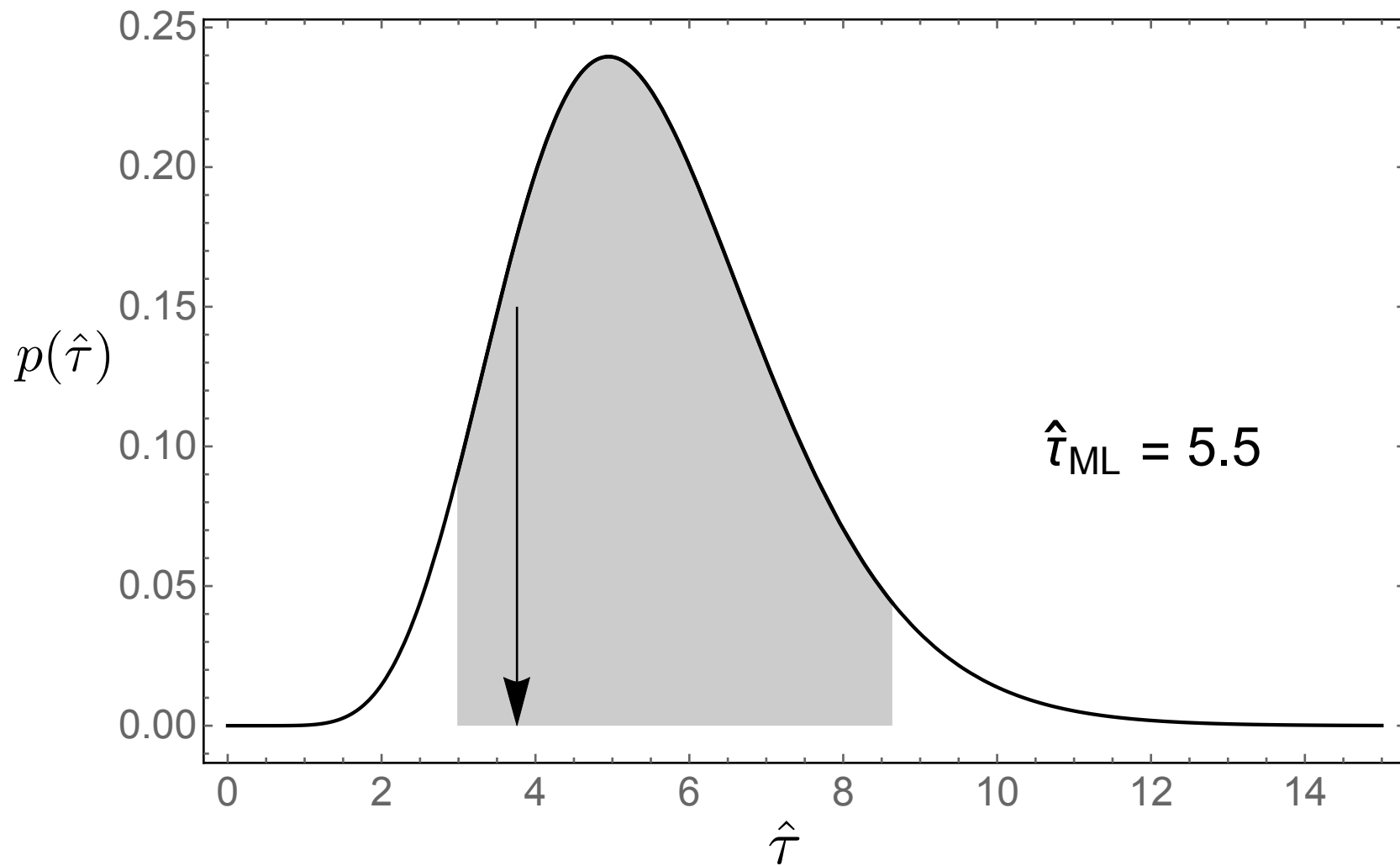
2. Central intervals as random variables

Different estimates of the decay time produce different central intervals



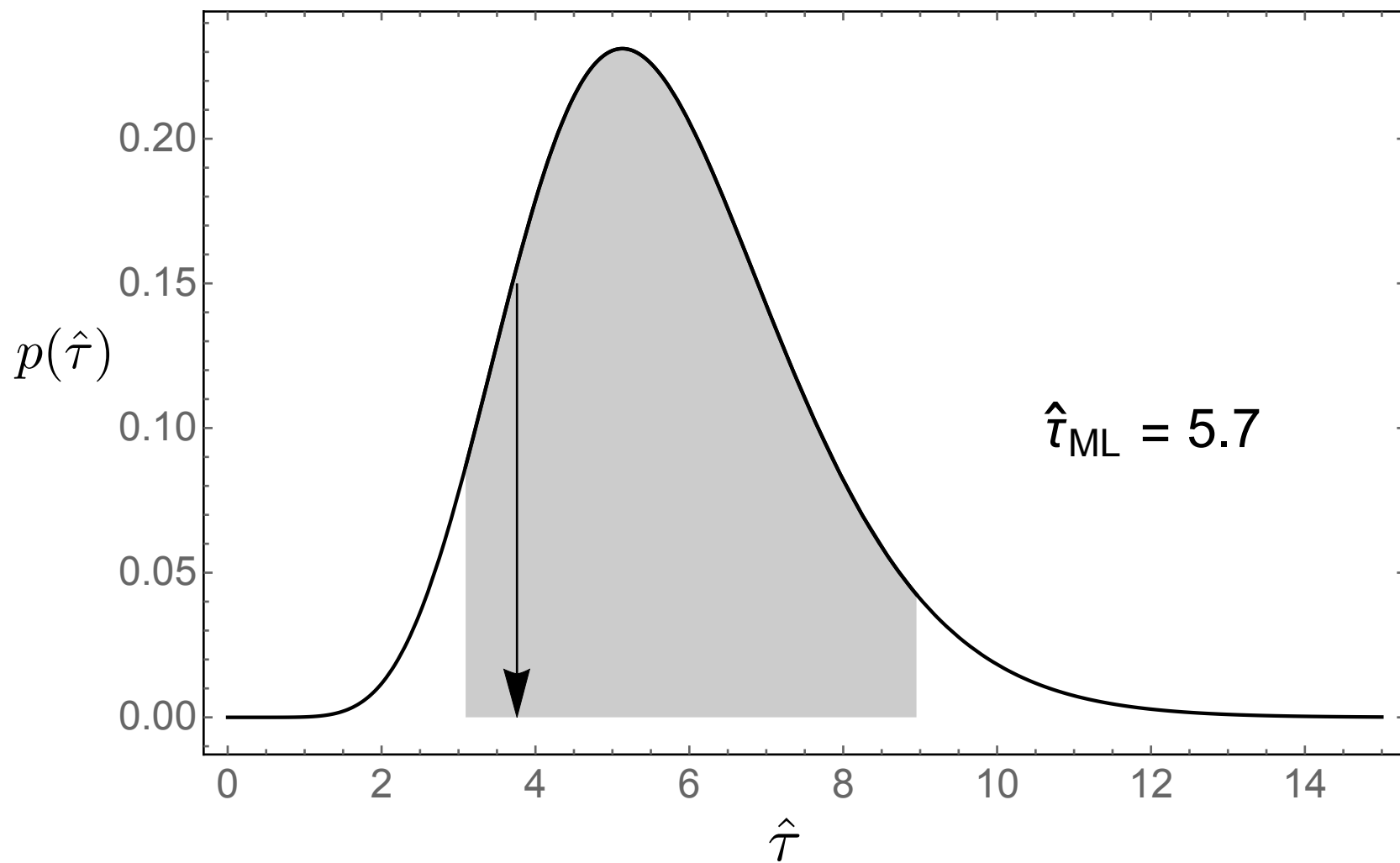
2. Central intervals as random variables

Different estimates of the decay time produce different central intervals



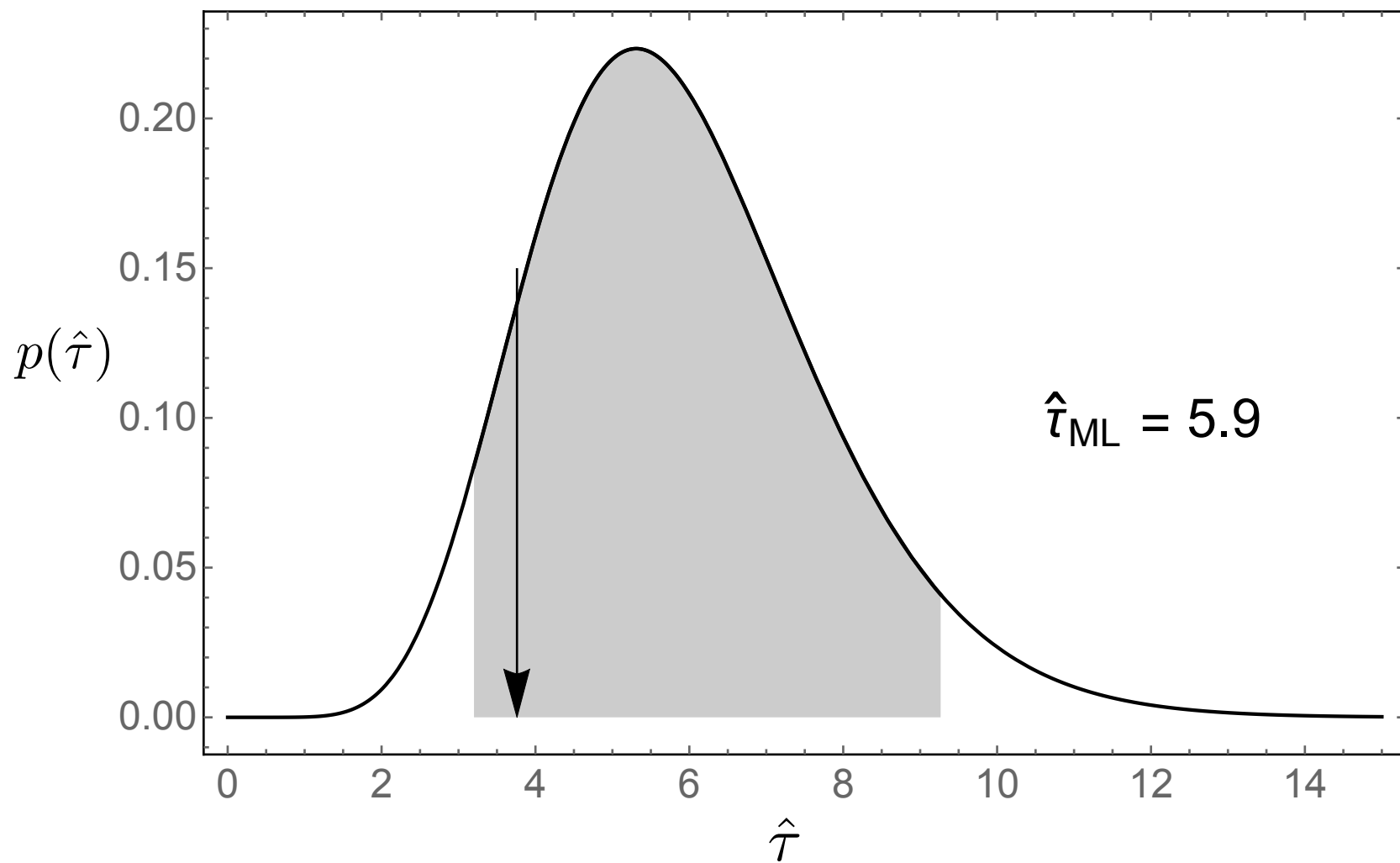
2. Central intervals as random variables

Different estimates of the decay time produce different central intervals



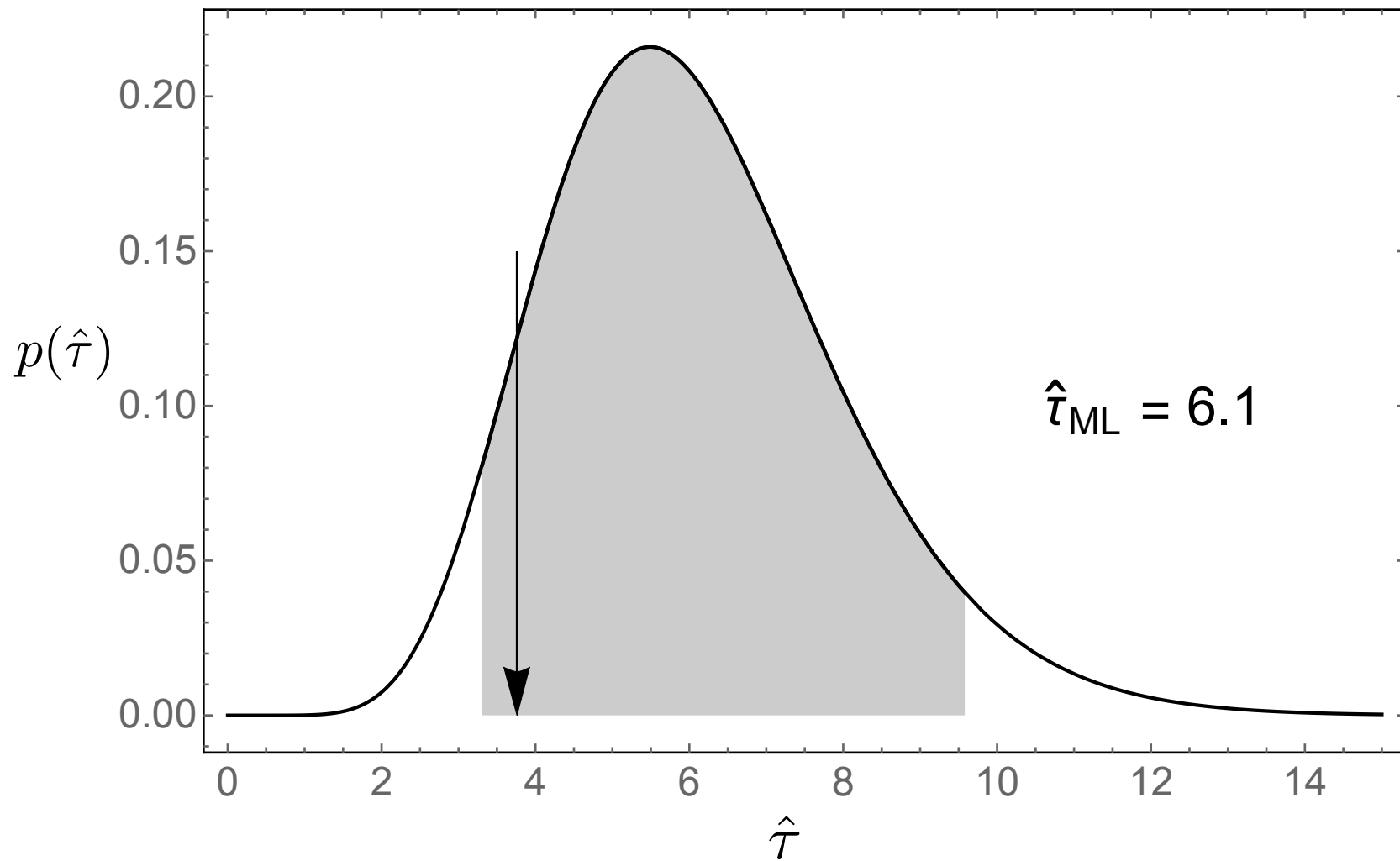
2. Central intervals as random variables

Different estimates of the decay time produce different central intervals



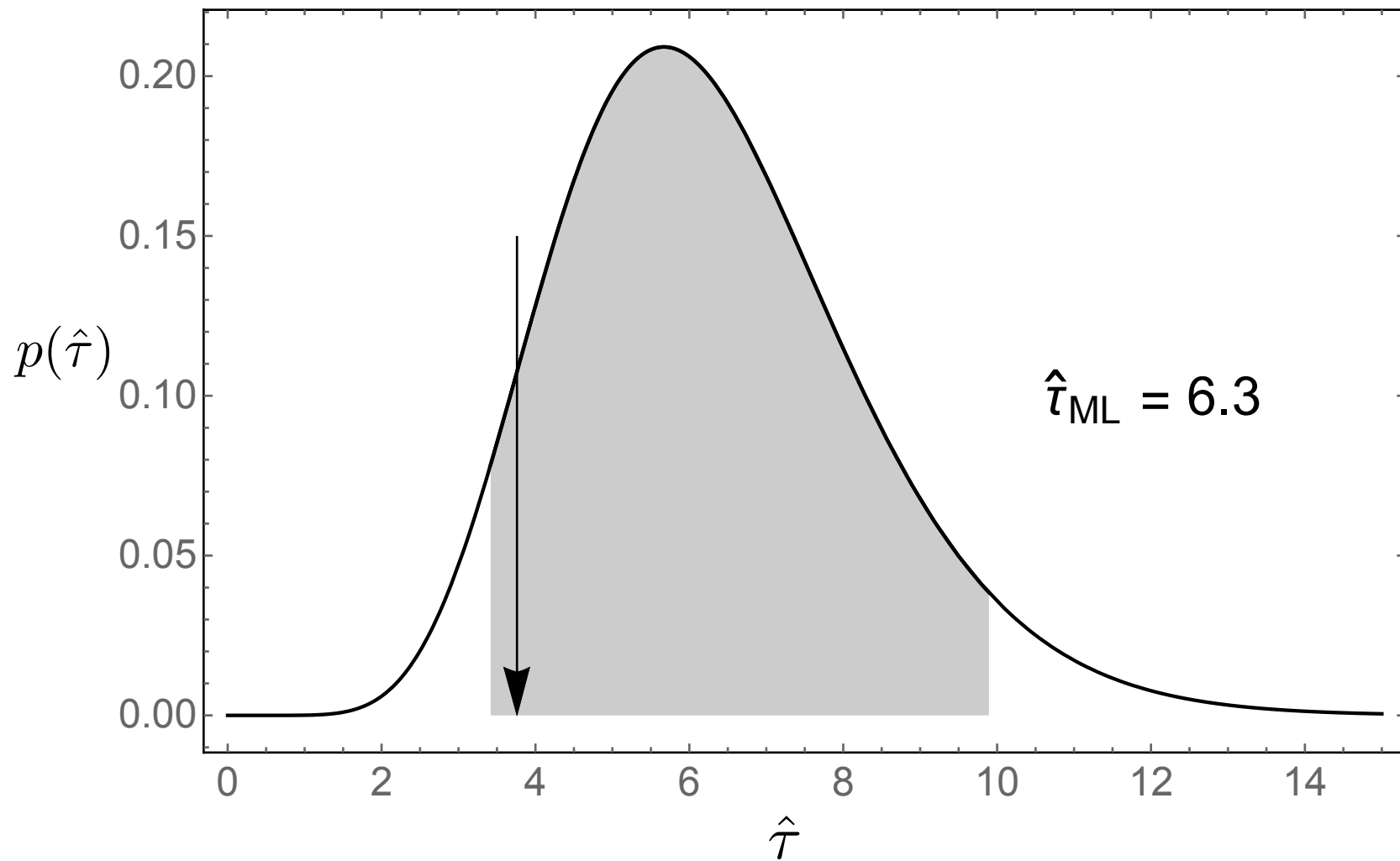
2. Central intervals as random variables

Different estimates of the decay time produce different central intervals



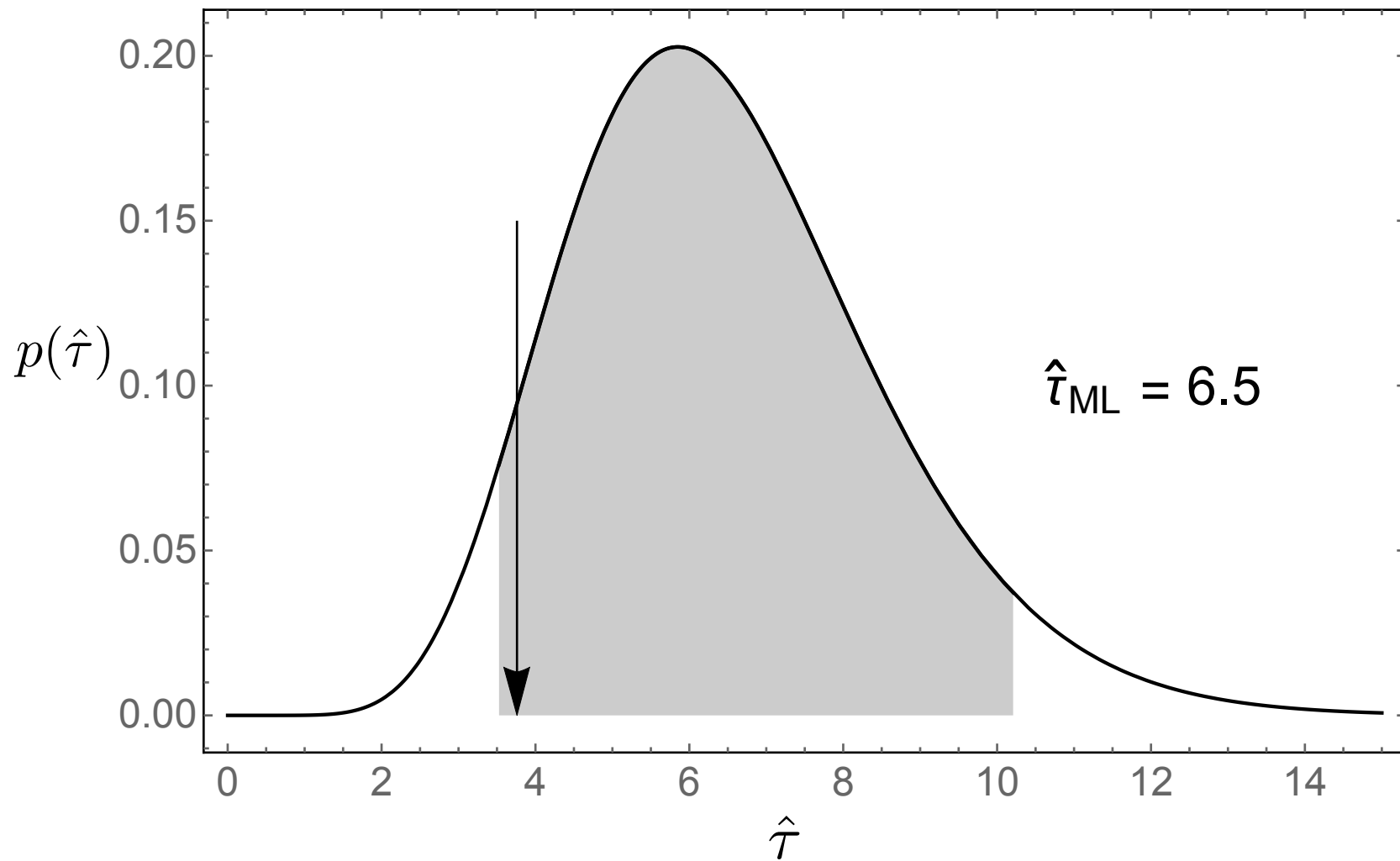
2. Central intervals as random variables

Different estimates of the decay time produce different central intervals



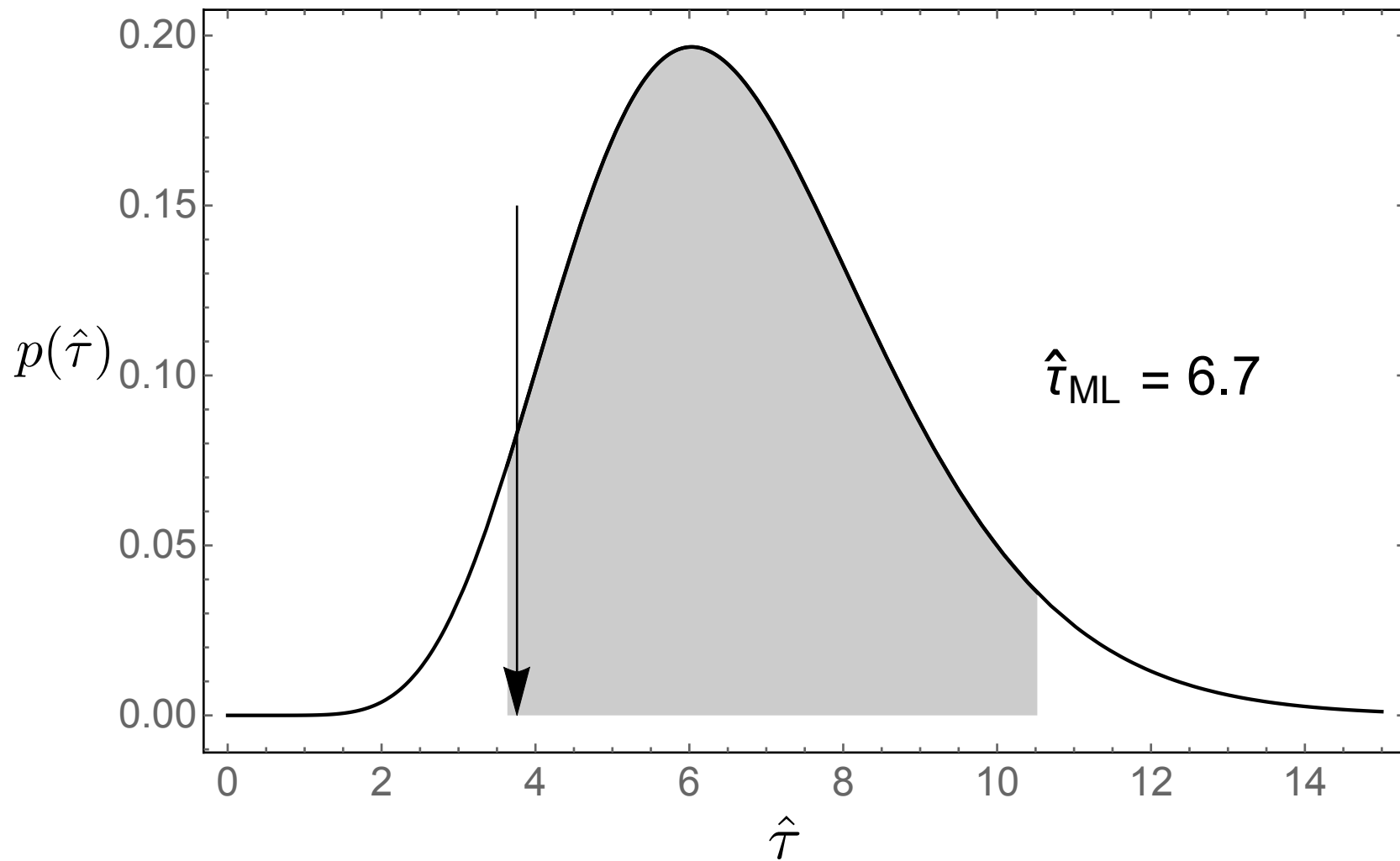
2. Central intervals as random variables

Different estimates of the decay time produce different central intervals



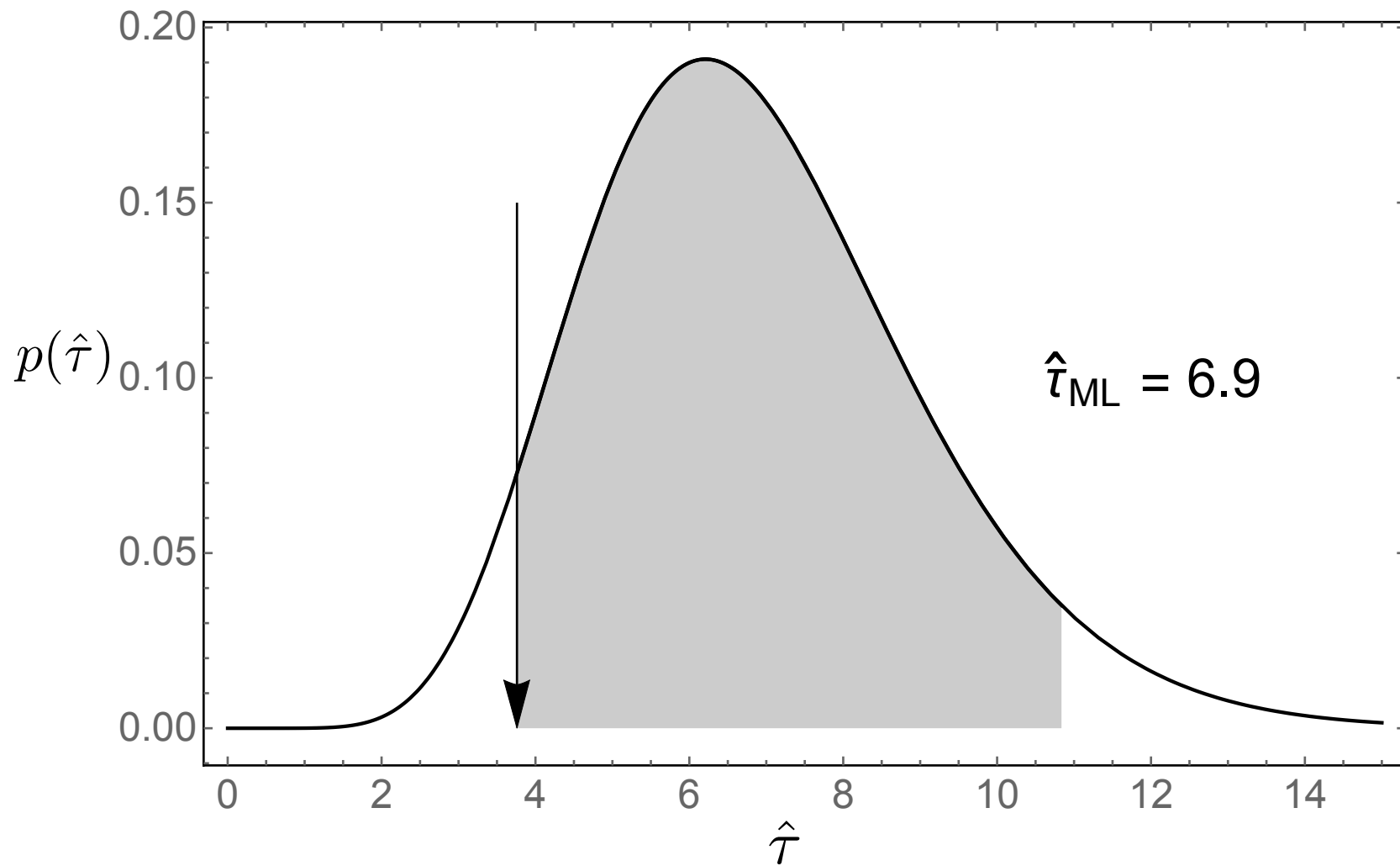
2. Central intervals as random variables

Different estimates of the decay time produce different central intervals



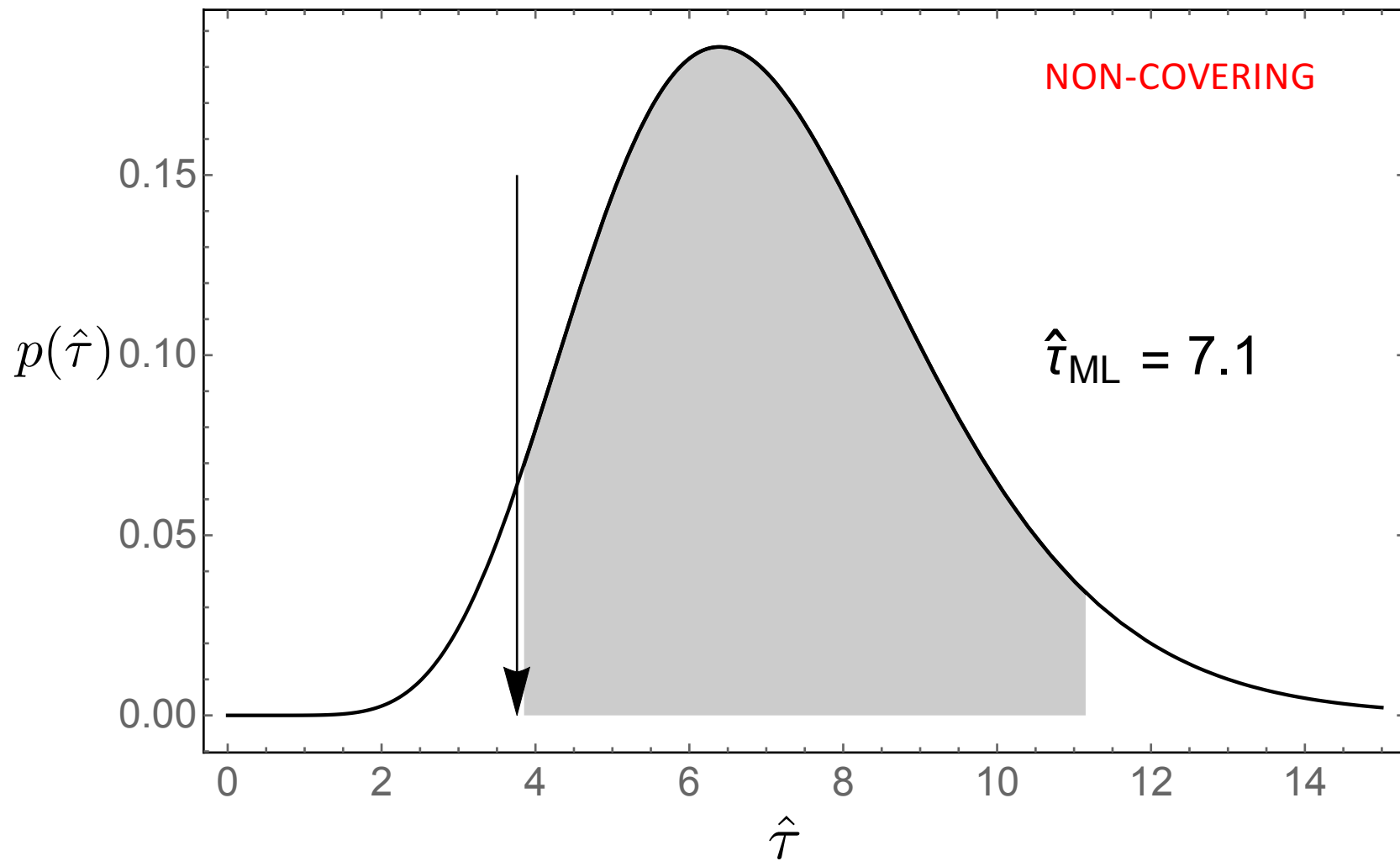
2. Central intervals as random variables

Different estimates of the decay time produce different central intervals



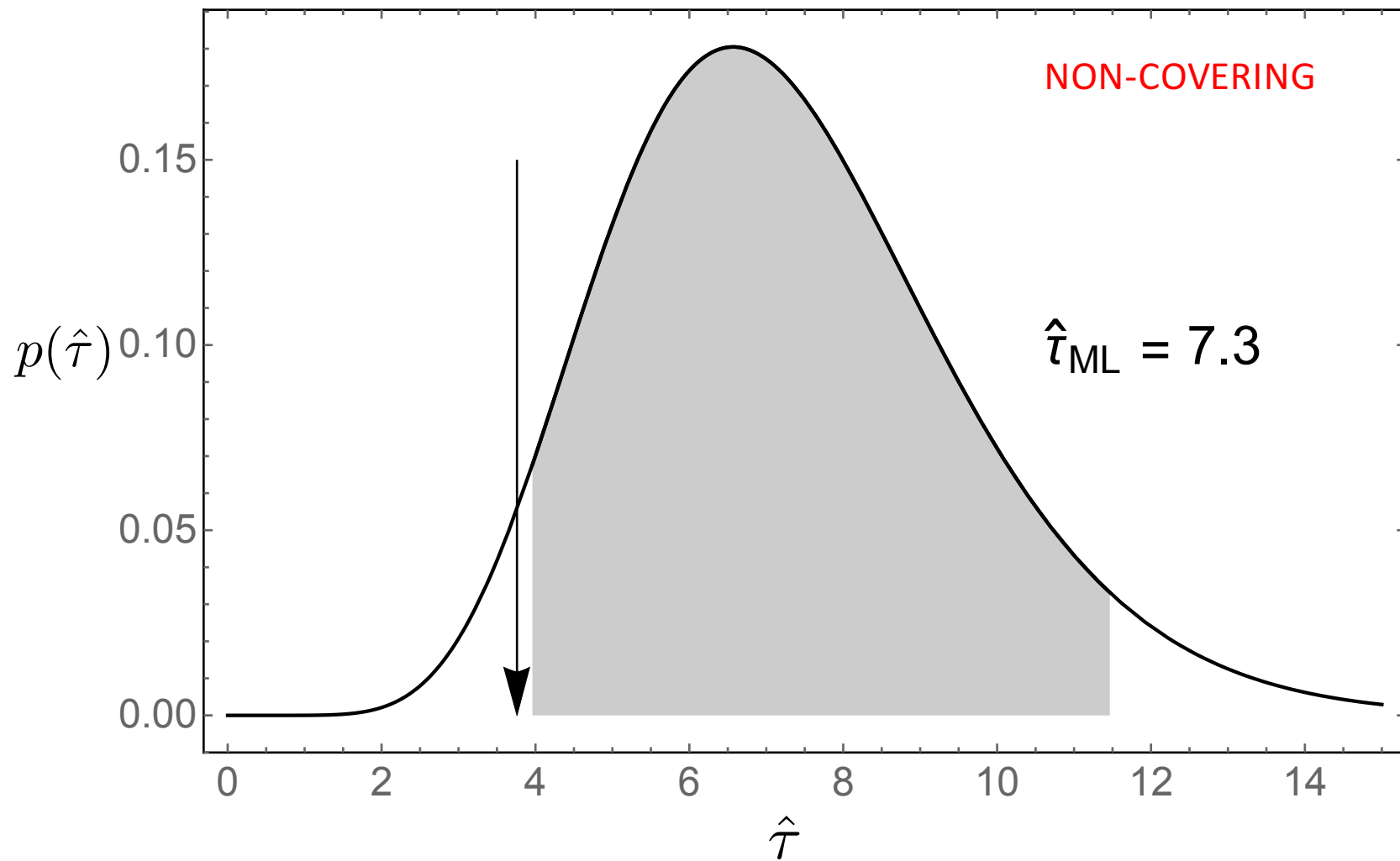
2. Central intervals as random variables

Different estimates of the decay time produce different central intervals



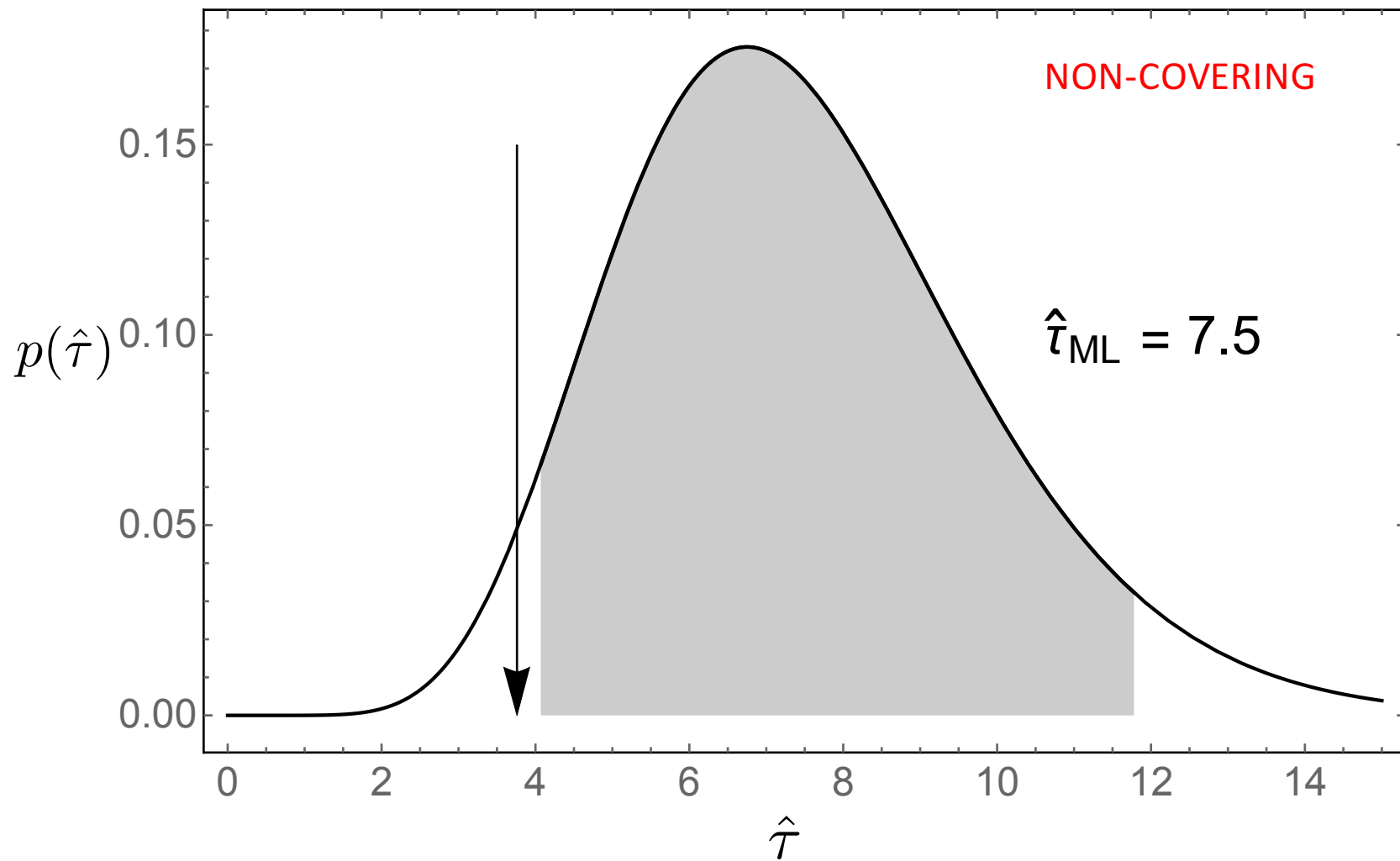
2. Central intervals as random variables

Different estimates of the decay time produce different central intervals



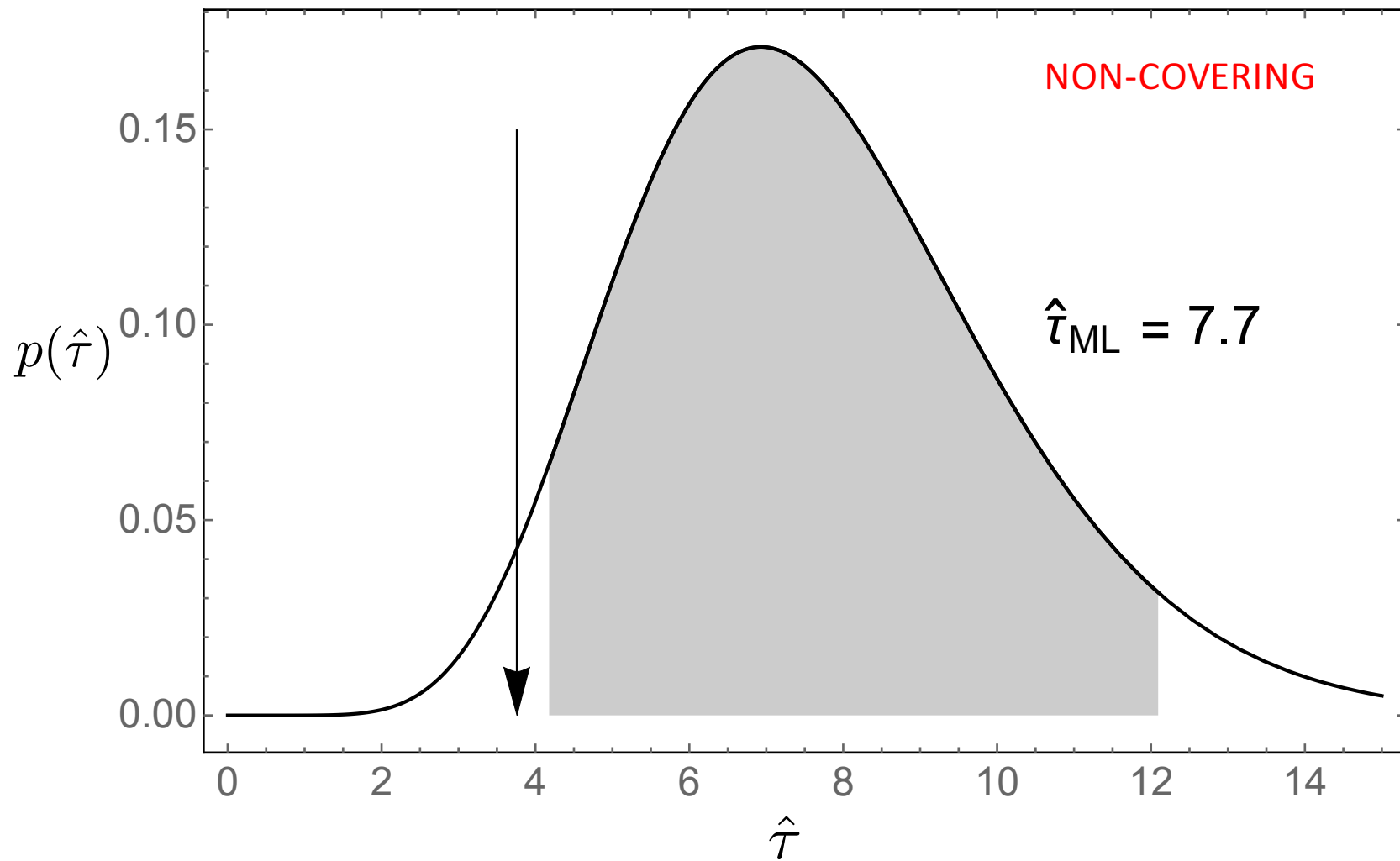
2. Central intervals as random variables

Different estimates of the decay time produce different central intervals



2. Central intervals as random variables

Different estimates of the decay time produce different central intervals

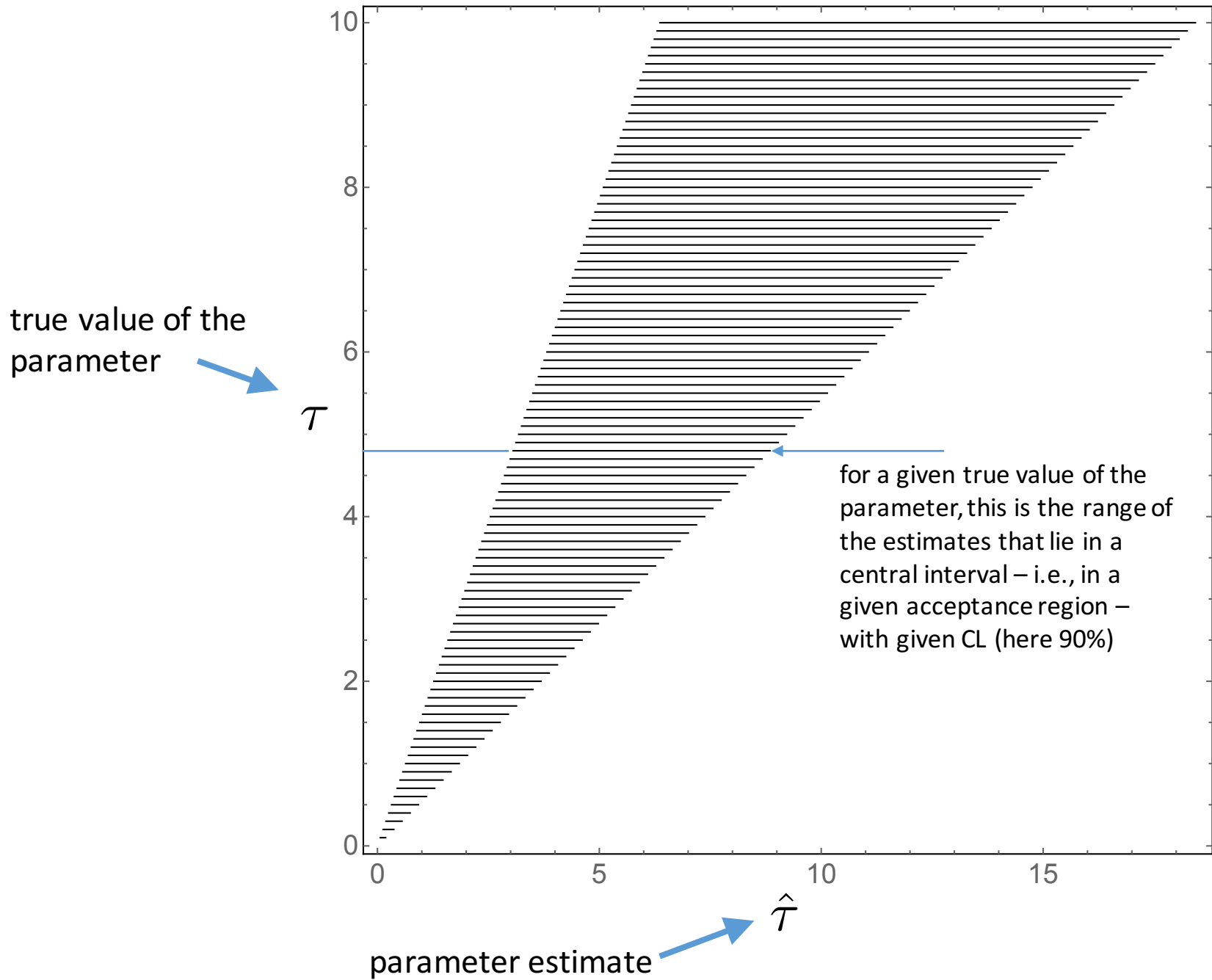


3. Range of estimates that produce a covering interval

In the example shown, all the values of the estimate $\hat{\tau}_{ML}$ in the interval between 2.39 and 6.93 have central intervals with confidence level = 90% that cover the true value of the parameter (3.76).

Unfortunately, we do not know the true value of the parameter ...

However we know how we could repeat the same construction over and over again for different true values of the parameter.

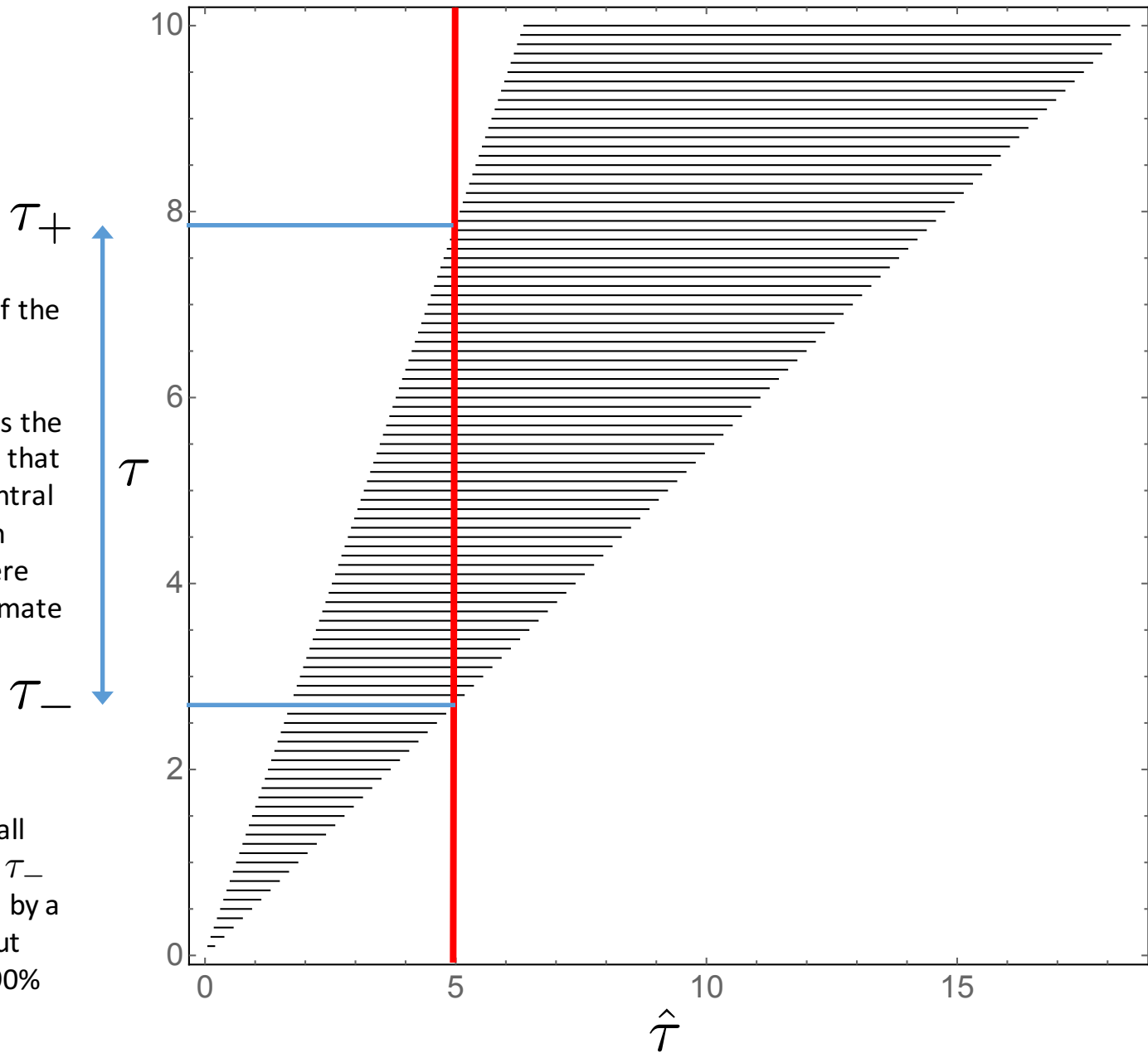


Corollary I—It follows from the Proposition I and II that whatever may be the true values $\theta'_1, \theta'_2 \dots \theta'_l$ of the θ 's, the probability of any fixed value θ''_1 of θ_1 being covered by $\delta(E)$ is identical with the probability of the sample point E falling within $A(\theta''_1)$.

$$\begin{aligned} P\{\delta(E) \subset \theta''_1 | \theta'_1, \dots, \theta'_l\} &= P\{\underline{\theta}(E) \leq \theta''_1 \leq \bar{\theta}(E) | \theta'_1, \theta'_2, \dots, \theta'_l\} \\ &= P\{E \in A(\theta''_1) | \theta'_1, \theta'_2, \dots, \theta'_l\}. \end{aligned} \quad (26)$$

Given an estimate of the parameter obtained from actual measurement, this is the range of true values that are covered by a central interval with a given confidence level (here 90%) about the estimate

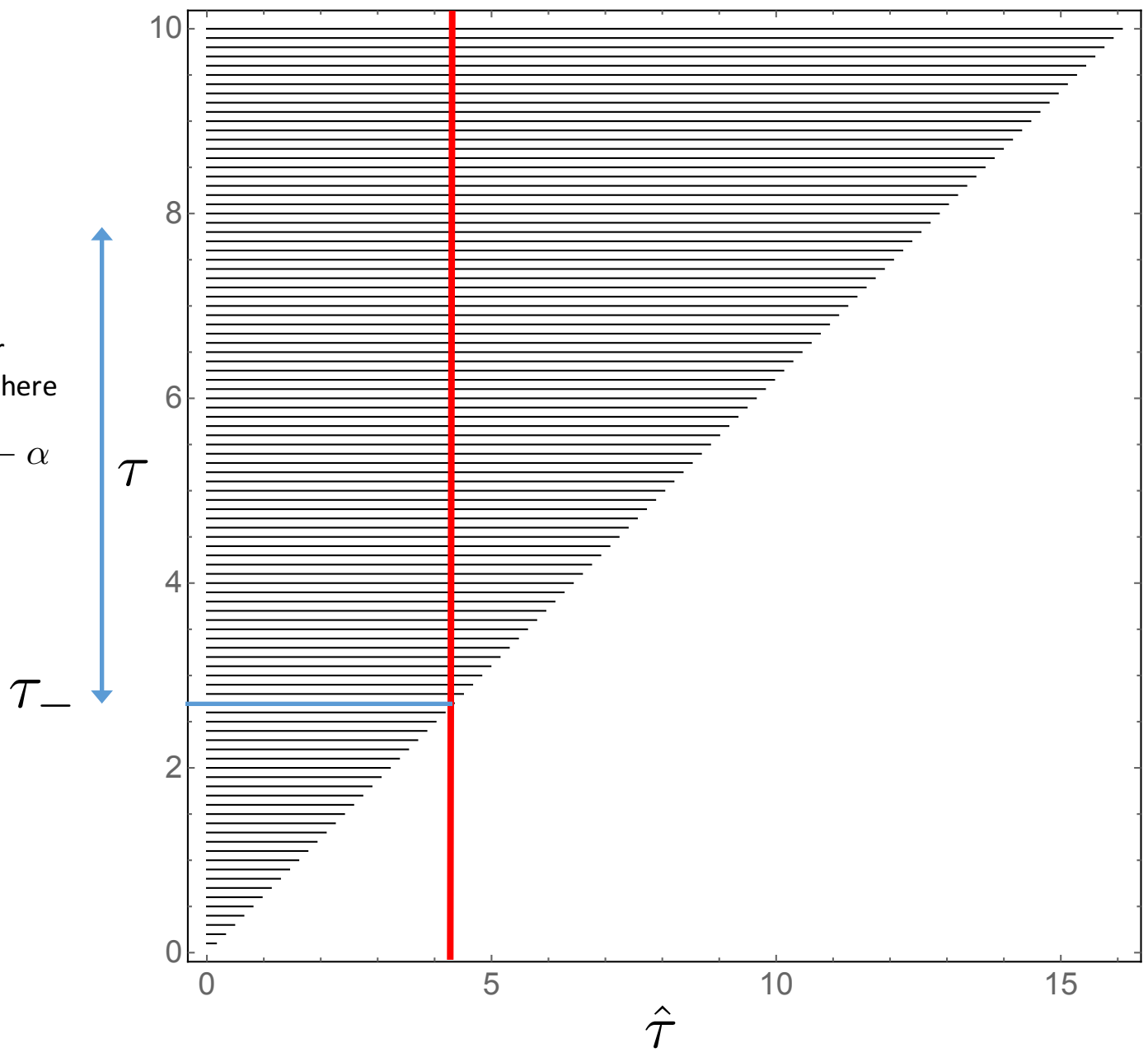
Here we know that all the values between τ_- and τ_+ are covered by a central interval about the estimate, with 90% probability.



This is instead an example of a “upper confidence limit”, where

$$P(\tau > \tau_-) = 1 - \alpha$$

(in this example $\alpha = 10\%$)



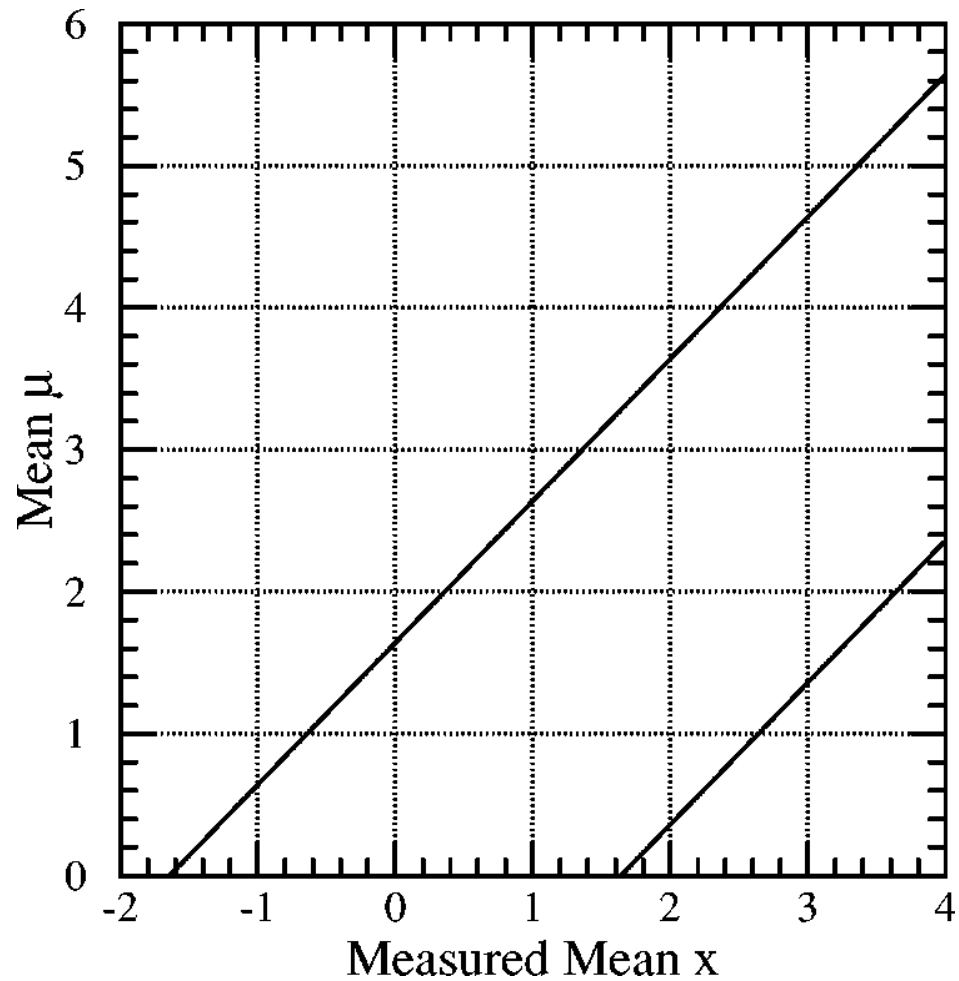


FIG. 3. Standard confidence belt for 90% C.L. central confidence intervals for the mean of a Gaussian, in units of the rms deviation.

5. Coverage

Intervals can overcover (probability more than confidence level) or undercover (probability less than confidence level). Neither option is OK, but overcovering better than undercovering ...

5. Confidence intervals and hypothesis testing

It is interesting to note that the choice of a confidence interval can be viewed as a hypothesis test.

Here the hypothesis is that the parameter has a given value, and one excludes all the values of the parameter that would be rejected with a given confidence level.

If this is the case, it is possible to choose a test statistics, and one obvious choice is to use the likelihood ratio:

$$t = \frac{L(D; \theta)}{L(D; \hat{\theta})}$$

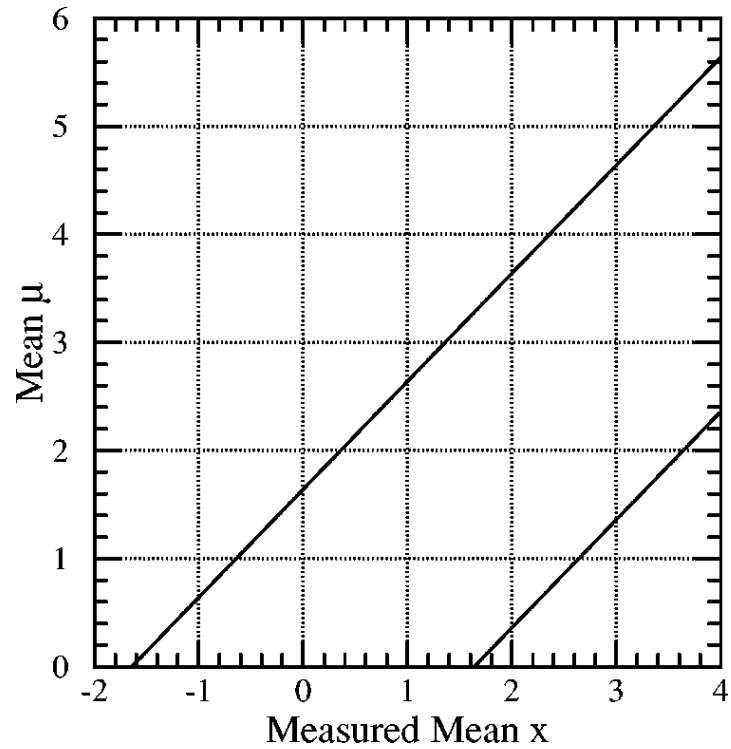
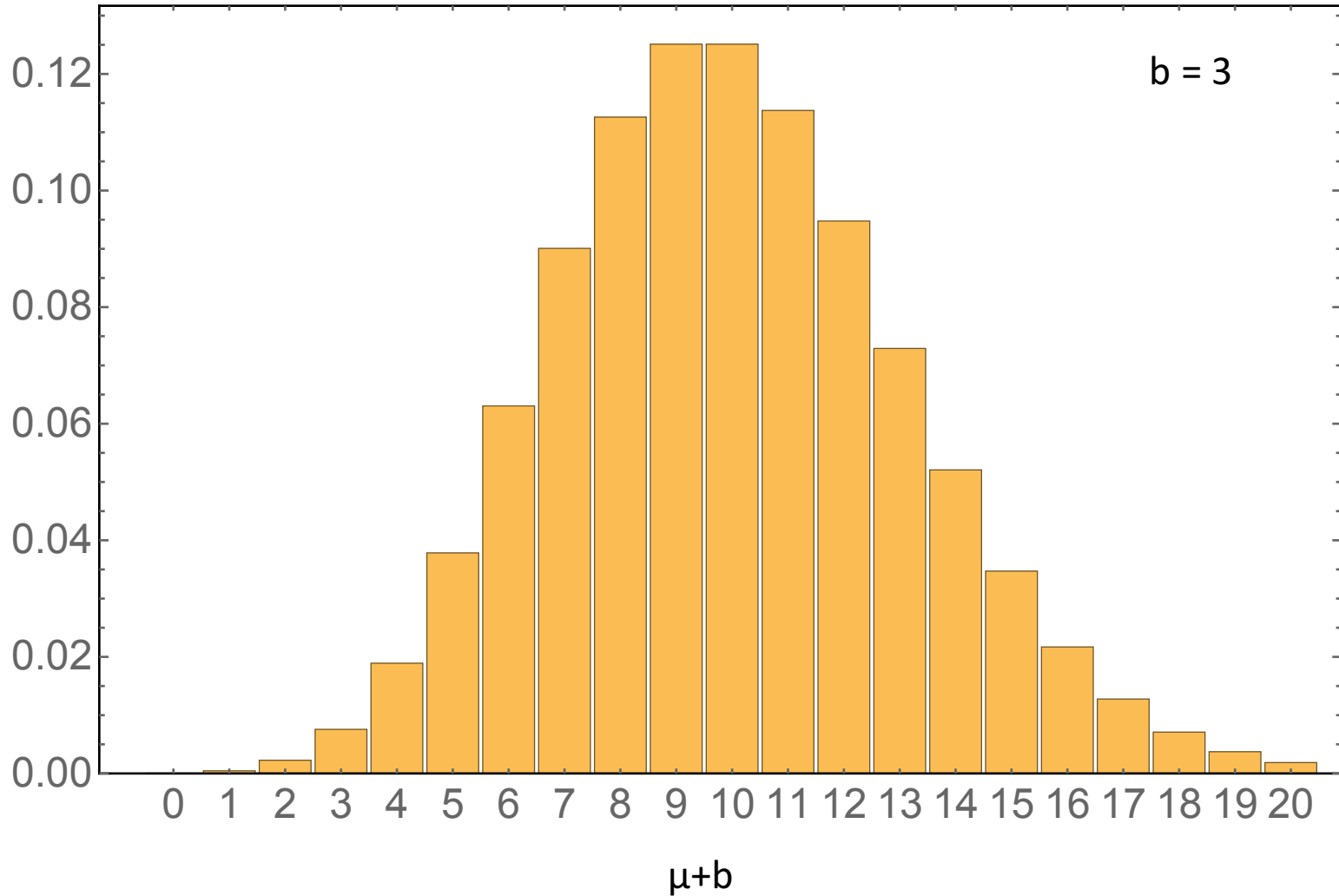


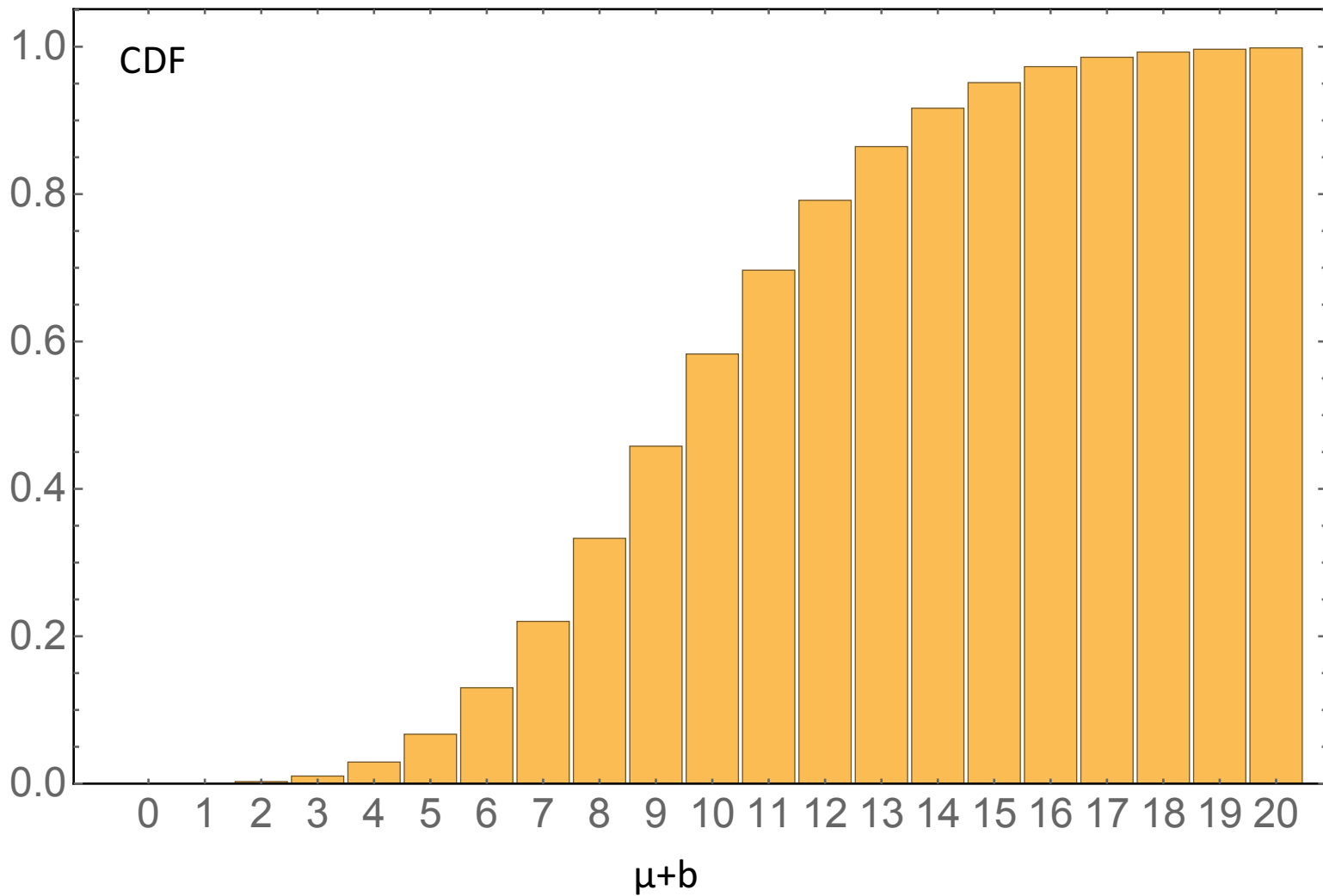
FIG. 3. Standard confidence belt for 90% C.L. central confidence intervals for the mean of a Gaussian, in units of the rms deviation.

Problem: the standard Gaussian confidence belt leads to negative values, and this is not always acceptable ...

Poisson process with known background

$$P(n|\mu) = \frac{(\mu + b)^n}{n!} e^{-(\mu+b)}$$





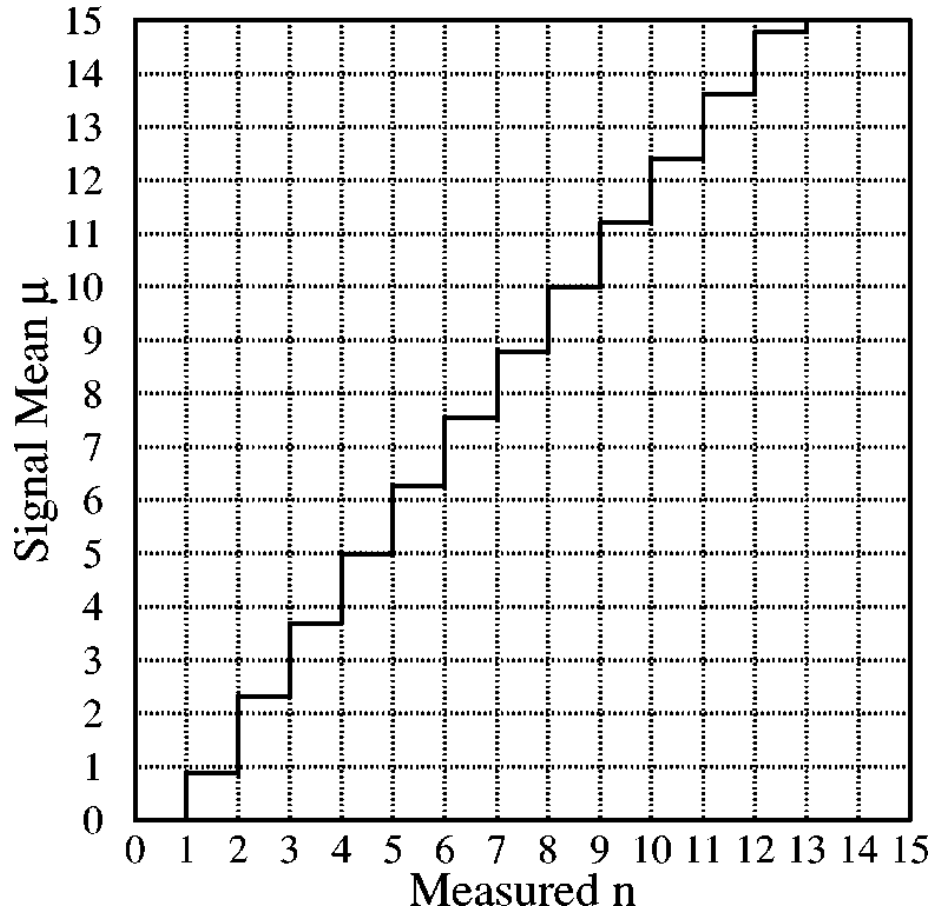
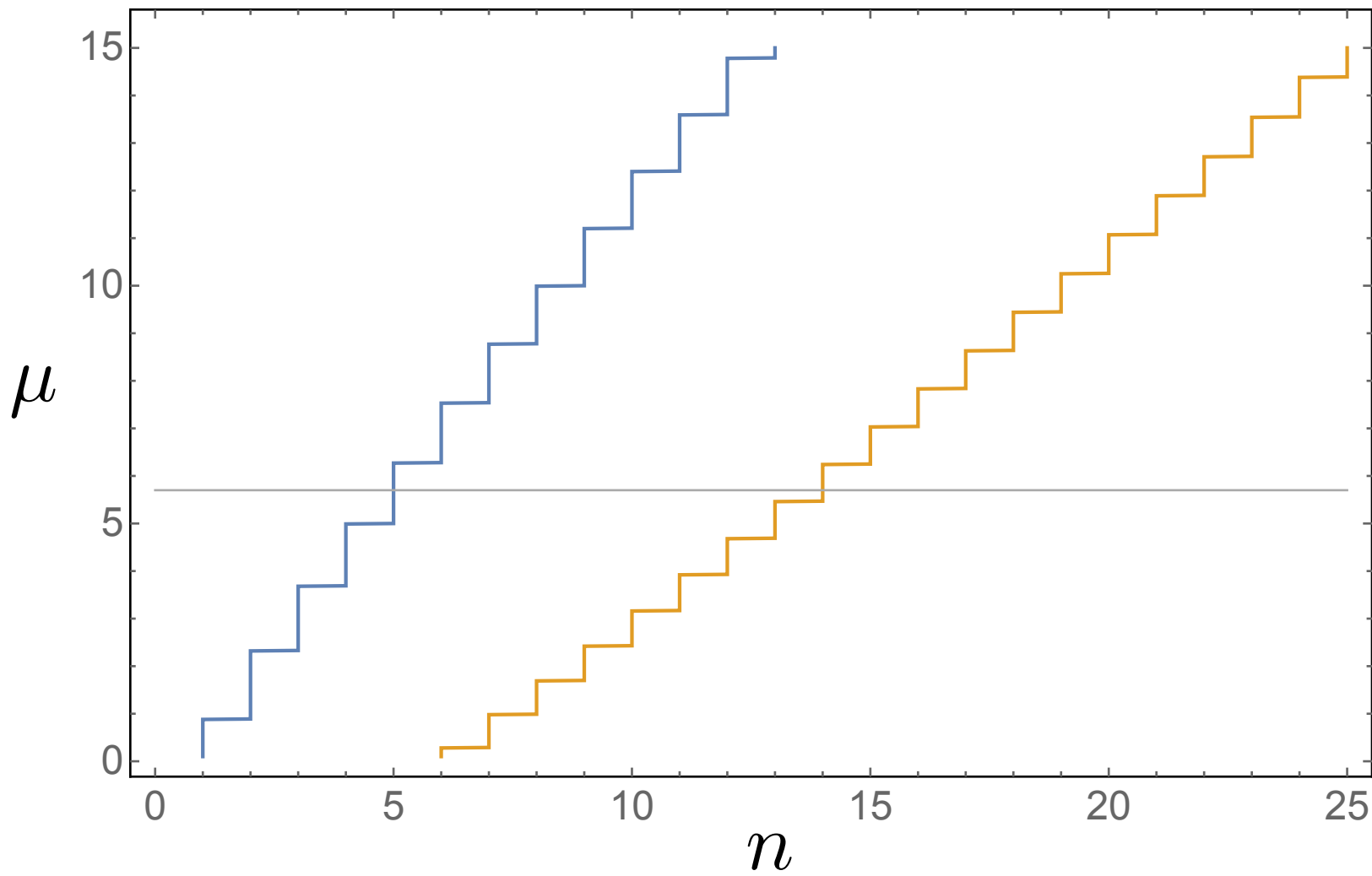
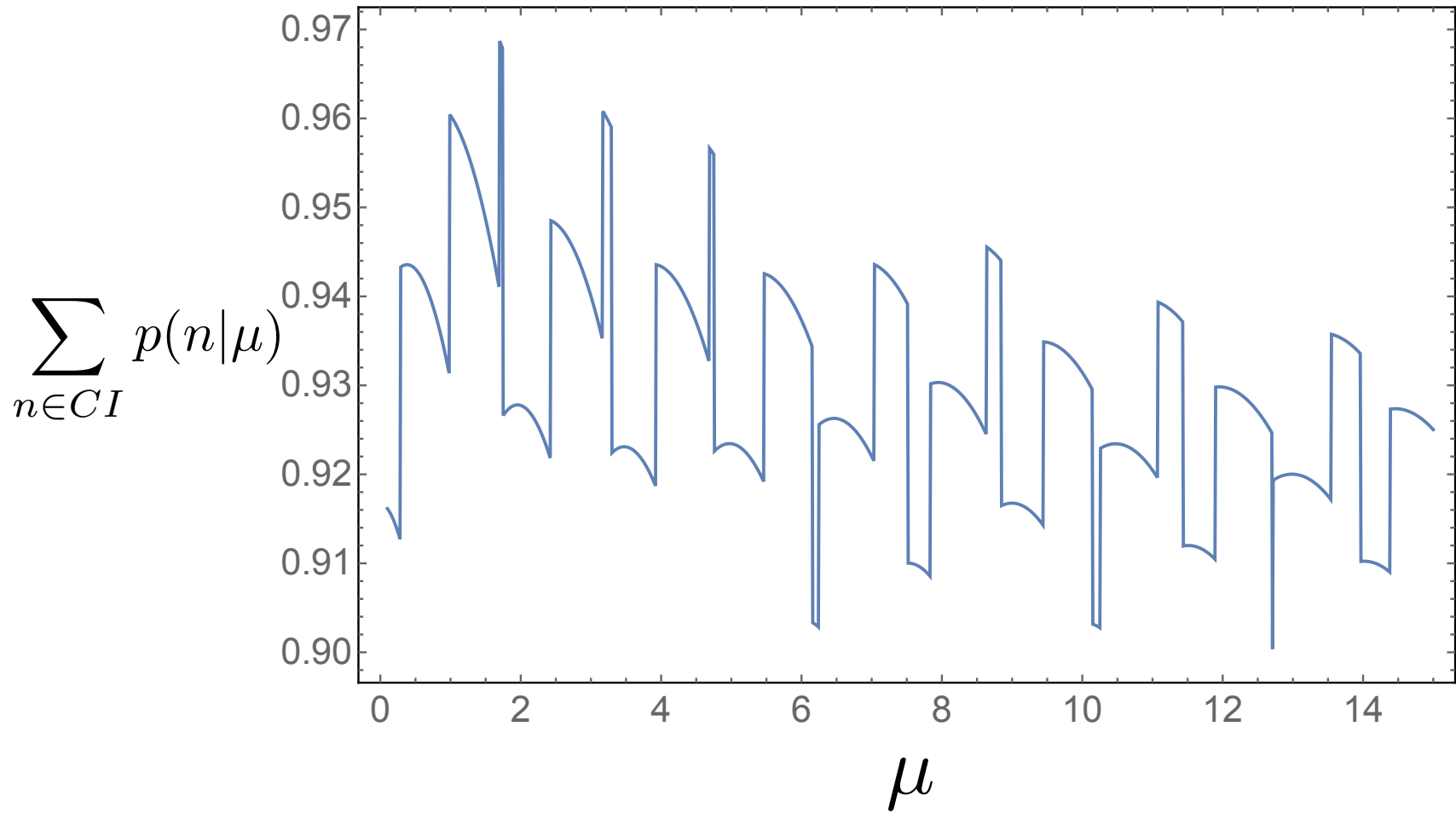


FIG. 5. Standard confidence belt for 90% C.L. upper limits, for unknown Poisson signal mean μ in the presence of a Poisson background with known mean $b=3.0$. The second line in the belt is at $n = +\infty$.



coverage



To avoid problems at small values use a different choice of CI's (Feldman & Cousins, 1998)

PHYSICAL REVIEW D

VOLUME 57, NUMBER 7

1 APRIL 1998

Unified approach to the classical statistical analysis of small signals

Gary J. Feldman^{*}

Department of Physics, Harvard University, Cambridge, Massachusetts 02138

Robert D. Cousins[†]

Department of Physics and Astronomy, University of California, Los Angeles, California 90095

(Received 21 November 1997; published 6 March 1998)

We give a classical confidence belt construction which unifies the treatment of upper confidence limits for null results and two-sided confidence intervals for non-null results. The unified treatment solves a problem (apparently not previously recognized) that the choice of upper limit or two-sided intervals leads to intervals which are not confidence intervals if the choice is based on the data. We apply the construction to two related problems which have recently been a battleground between classical and Bayesian statistics: Poisson processes with background and Gaussian errors with a bounded physical region. In contrast with the usual classical construction for upper limits, our construction avoids unphysical confidence intervals. In contrast with some popular Bayesian intervals, our intervals eliminate conservatism (frequentist coverage greater than the stated confidence) in the Gaussian case and reduce it to a level dictated by discreteness in the Poisson case. We generalize the method in order to apply it to analysis of experiments searching for neutrino oscillations. We show that this technique both gives correct coverage and is powerful, while other classical techniques that have been used by neutrino oscillation search experiments fail one or both of these criteria.

[S0556-2821(98)00109-X]

PACS number(s): 06.20.Dk, 14.60.Pq

We begin with a numerical example which occurs in the construction of confidence belts for a Poisson process with background. The construction proceeds in the manner of Fig. 1, where the measurement x in Fig. 1 now corresponds to the measured total number of events n .

Let the known mean background be $b = 3.0$, and consider the construction of the horizontal acceptance interval at signal mean $\mu = 0.5$. Then $P(n|\mu)$ is given by Eq. (3.2), and is given in the second column of Table I.

Now consider, for example, $n = 0$. For the assumed $b = 3.$, the probability of obtaining 0 events is 0.03 if $\mu = 0.5$, which is quite low on an absolute scale. However, it is not so low when compared to the probability (0.05) of obtaining 0 events with $b = 3.0$ and $\mu = 0.0$, which is the alternate hypothesis with the greatest likelihood. A *ratio* of likelihoods, in this case 0.03/0.05, is what we use as our ordering principle when selecting those values of n to place in the acceptance interval.

That is, for each n , we let μ_{best} be that value of the mean signal μ which maximizes $P(n|\mu)$; we require μ_{best} to be physically allowed, i.e., non-negative in this case. Then $\mu_{\text{best}} = \max(0, n - b)$, and is given in the third column of Table I. We then compute $P(n|\mu_{\text{best}})$, which is given in the fourth column. The fifth column contains the ratio

$$R = P(n|\mu) / P(n|\mu_{\text{best}}), \quad (4.1)$$

and is the quantity on which our ordering principle is based. R is a ratio of two likelihoods: the likelihood of obtaining n given the actual mean μ , and the likelihood of obtaining n given the best-fit physically allowed mean.

given the best-fit physically allowed mean. Values of n are added to the acceptance region for a given μ in decreasing order of R , until the sum of $P(n|\mu)$ meets or exceeds the desired C.L. This ordering, for values of n necessary to obtain total probability of 90%, is shown in the column labeled “rank.” Thus, the acceptance region for $\mu = 0.5$ is the interval $n = [0,6]$. Because of the discreteness of n , the acceptance region contains a summed probability greater than 90%; this is unavoidable no matter what the ordering principle, and leads to confidence intervals which are conservative.

TABLE I. Illustrative calculations in the confidence belt construction for signal mean μ in the presence of known mean background $b=3.0$. Here we find the acceptance interval for $\mu=0.5$.

n	$P(n \mu)$	μ_{best}	$P(n \mu_{\text{best}})$	R	rank	U.L.	central
0	0.030	0.0	0.050	0.607	6		
1	0.106	0.0	0.149	0.708	5	✓	✓
2	0.185	0.0	0.224	0.826	3	✓	✓
3	0.216	0.0	0.224	0.963	2	✓	✓
4	0.189	1.0	0.195	0.966	1	✓	✓
5	0.132	2.0	0.175	0.753	4	✓	✓
6	0.077	3.0	0.161	0.480	7	✓	✓
7	0.039	4.0	0.149	0.259		✓	✓
8	0.017	5.0	0.140	0.121		✓	
9	0.007	6.0	0.132	0.050		✓	
10	0.002	7.0	0.125	0.018		✓	
11	0.001	8.0	0.119	0.006		✓	

With a Gaussian distribution, things proceed similarly, with the difference that now one takes non-central intervals ...

$$P(x|\mu_{\text{best}}) = \begin{cases} 1/\sqrt{2\pi}, & x \geq 0, \\ \exp(-x^2/2)/\sqrt{2\pi}, & x < 0. \end{cases} \quad (4.2)$$

We then compute R in analogy to Eq. (4.1), using Eqs. (3.1) and (4.2):

$$R(x) = \frac{P(x|\mu)}{P(x|\mu_{\text{best}})} = \begin{cases} \exp(-(x-\mu)^2/2), & x \geq 0 \\ \exp(x\mu - \mu^2/2), & x < 0. \end{cases} \quad (4.3)$$

During our Neyman construction of confidence intervals, R determines the order in which values of x are added to the acceptance region at a particular value of μ . In practice, this means that for a given value of μ , one finds the interval $[x_1, x_2]$ such that $R(x_1) = R(x_2)$ and

$$\int_{x_1}^{x_2} P(x|\mu) dx = \alpha.$$

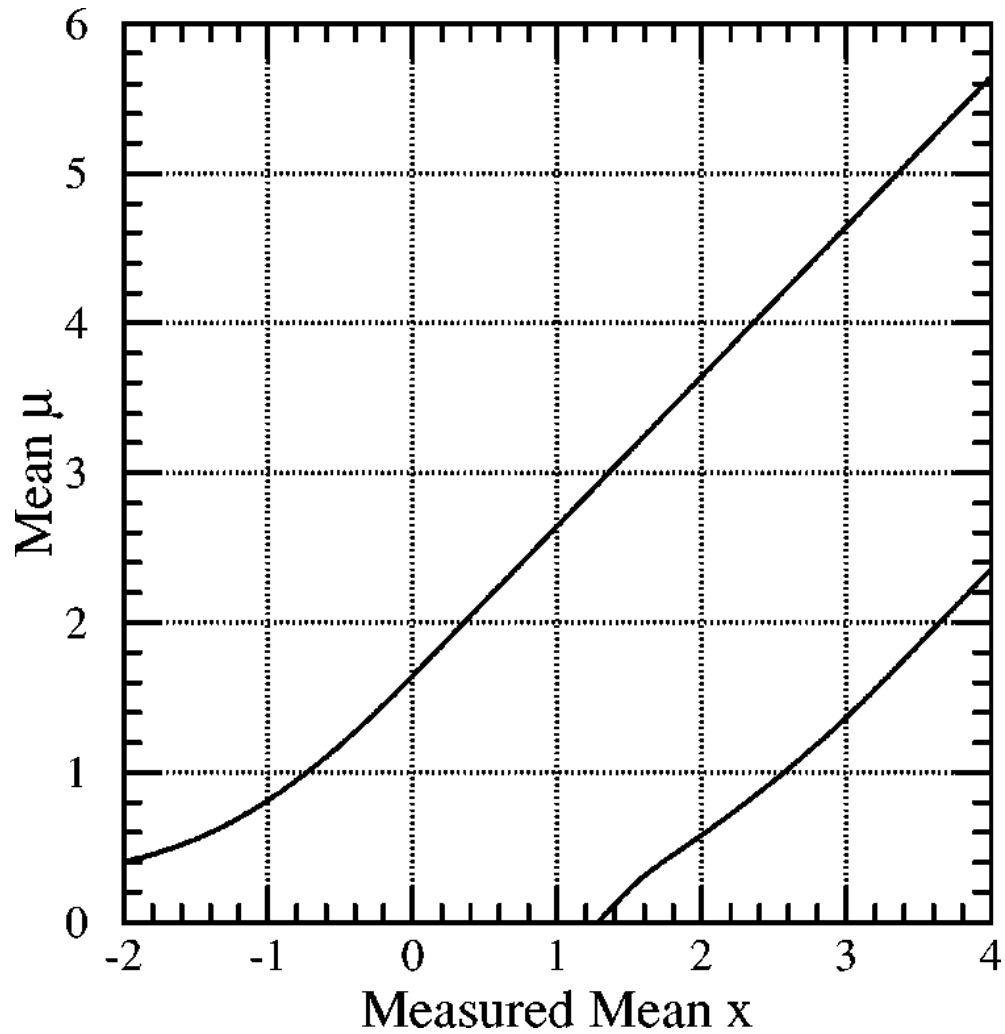


FIG. 10. Plot of our 90% confidence intervals for the mean of a Gaussian, constrained to be non-negative, described in the text.

The CLs method

The CLs method is based on a simple test statistic

$$CL_s = \frac{CL_{s+b}}{CL_b}$$

where

$$CL_{s+b} = P_{s+b}(X < X_{obs}); \quad CL_b = P_b(X < X_{obs})$$