90% Confidence Level Upper Bound

Brief discussion of Feldman & Cousins

10-31-2024

Outline

Parameter Estimataion

Poisson Distribution

- Expect an average of (real) $\mu \geq 0$ neutrinos per time

Poisson Distribution

- Expect an average of (real) $\mu \geq 0$ neutrinos per time
- Probability of seeing (integer) $n \geq 0$ neutrinos per time?

Poisson Distribution

- Expect an average of (real) $\mu \geq 0$ neutrinos per time
- Probability of seeing (integer) $n \geq 0$ neutrinos per time?
- Poisson Distribution:

$$\Pr(n|\mu) = \frac{e^{-\mu}\mu^n}{n!}$$

- Prob. of data n, given parameter μ; aka the *likelihood*.

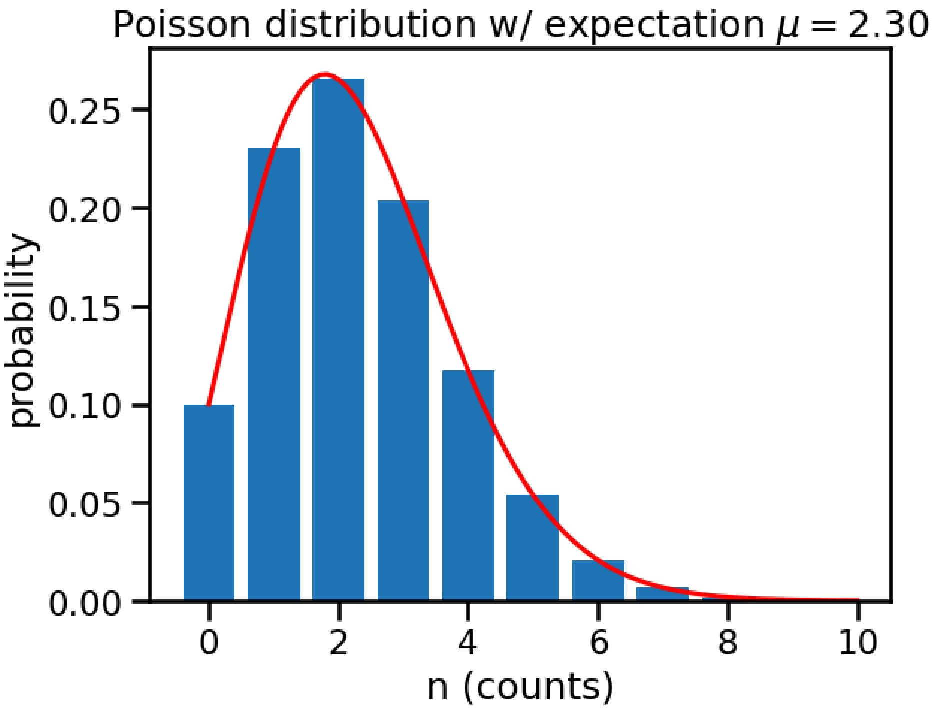
Poisson Distribution

- Expect an average of (real) $\mu \geq 0$ neutrinos per time
- Probability of seeing (integer) $n \geq 0$ neutrinos per time?
- Poisson Distribution:

$$\Pr(n|\mu) = \frac{e^{-\mu}\mu^n}{n!}$$

- Prob. of data n, given parameter μ; aka the *likelihood*.

- Example:



Confidence Interval (CI)

CI Definition

- Goal: estimate parameter μ whose true value is μ_t

CI Definition

- Goal: estimate parameter μ whose true value is μ_t
- Make a measurement x; first measurement yields x_0.

CI Definition

- Goal: estimate parameter μ whose true value is μ_t
- Make a measurement x; first measurement yields x_0.
- Construct an interval (discussed later) $[\mu_l, \mu_u]$

CI Definition

- Goal: estimate parameter μ whose true value is μ_t
- Make a measurement x; first measurement yields x_0.
- Construct an interval (discussed later) $[\mu_l, \mu_u]$
  - $\mu_l = \mu_l(x_0)$: lower bound associated w/ this 1^{st} measurement

CI Definition

- Goal: estimate parameter μ whose true value is μ_t
- Make a measurement x; first measurement yields x_0.
- Construct an interval (discussed later) $[\mu_l, \mu_u]$
  - $\mu_l = \mu_l(x_0)$: lower bound associated w/ this 1$^{\text{st}}$ measurement
  - $\mu_u = \mu_u(x_0)$: upper bound associated w/ this 1$^{\text{st}}$ measurement

CI Definition

- Goal: estimate parameter μ whose true value is μ_t
- Make a measurement x; first measurement yields x_0.
- Construct an interval (discussed later) $[\mu_l, \mu_u]$
  - $\mu_l = \mu_l(x_0)$: lower bound associated w/ this 1st measurement
  - $\mu_u = \mu_u(x_0)$: upper bound associated w/ this 1st measurement
- Repeat experiment; get outcome $x_1 \longrightarrow$ construct $[\mu_l(x_1), \mu_u(x_1)]$

CI Definition

- Goal: estimate parameter μ whose true value is μ_t
- Make a measurement x; first measurement yields x_0.
- Construct an interval (discussed later) $[\mu_l, \mu_u]$
  - $\mu_l = \mu_l(x_0)$: lower bound associated w/ this 1^{st} measurement
  - $\mu_u = \mu_u(x_0)$: upper bound associated w/ this 1^{st} measurement
- Repeat experiment; get outcome $x_1 \longrightarrow$ construct $[\mu_l(x_1), \mu_u(x_1)]$
- More experiments; get a bunch of intervals. *i.e.* we get a set

$$C \equiv \{[\mu_l(x_0), \mu_u(x_0)], [\mu_l(x_1), \mu_u(x_1)], [\mu_l(x_2), \mu_u(x_2)]...\}$$

CI Definition

- $C \equiv \{[\mu_l, \mu_u], [\mu_l, \mu_u], [\mu_l, \mu_u]...\}$
- The set C has the property that

$$P([\mu_l, \mu_u] \ni \mu_t) = \alpha\%$$

CI Definition

- $C \equiv \{[\mu_l, \mu_u], [\mu_l, \mu_u], [\mu_l, \mu_u]...\}$
- The set C has the property that

$$P([\mu_l, \mu_u] \ni \mu_t) = \alpha\%$$

- In words:
  - the true value μ_t is *fixed* (although unknown)

CI Definition

- $C \equiv \{[\mu_l, \mu_u], [\mu_l, \mu_u], [\mu_l, \mu_u]...\}$
- The set C has the property that

$$P([\mu_l, \mu_u] \ni \mu_t) = \alpha\%$$

- In words:
  - the true value μ_t is *fixed* (although unknown)
  - Some members of C would *cover* this true value

CI Definition

- $C \equiv \{[\mu_l, \mu_u], [\mu_l, \mu_u], [\mu_l, \mu_u] \ldots\}$
- The set C has the property that

$$P([\mu_l, \mu_u] \ni \mu_t) = \alpha\%$$

- In words:
  - the true value μ_t is *fixed* (although unknown)
  - Some members of C would *cover* this true value
  - e.g. say $\mu_t = 2$, and maybe in C there is $[1, 3]$.

CI Definition

- $C \equiv \{[\mu_l, \mu_u], [\mu_l, \mu_u], [\mu_l, \mu_u]...\}$
- The set C has the property that

$$P([\mu_l, \mu_u] \ni \mu_t) = \alpha\%$$

- In words:
  - the true value μ_t is *fixed* (although unknown)
  - Some members of C would *cover* this true value
  - e.g. say $\mu_t = 2$, and maybe in C there is $[1, 3]$.
- The members of C are called *confidence intervals*.

CI Construction: Confidence Belt

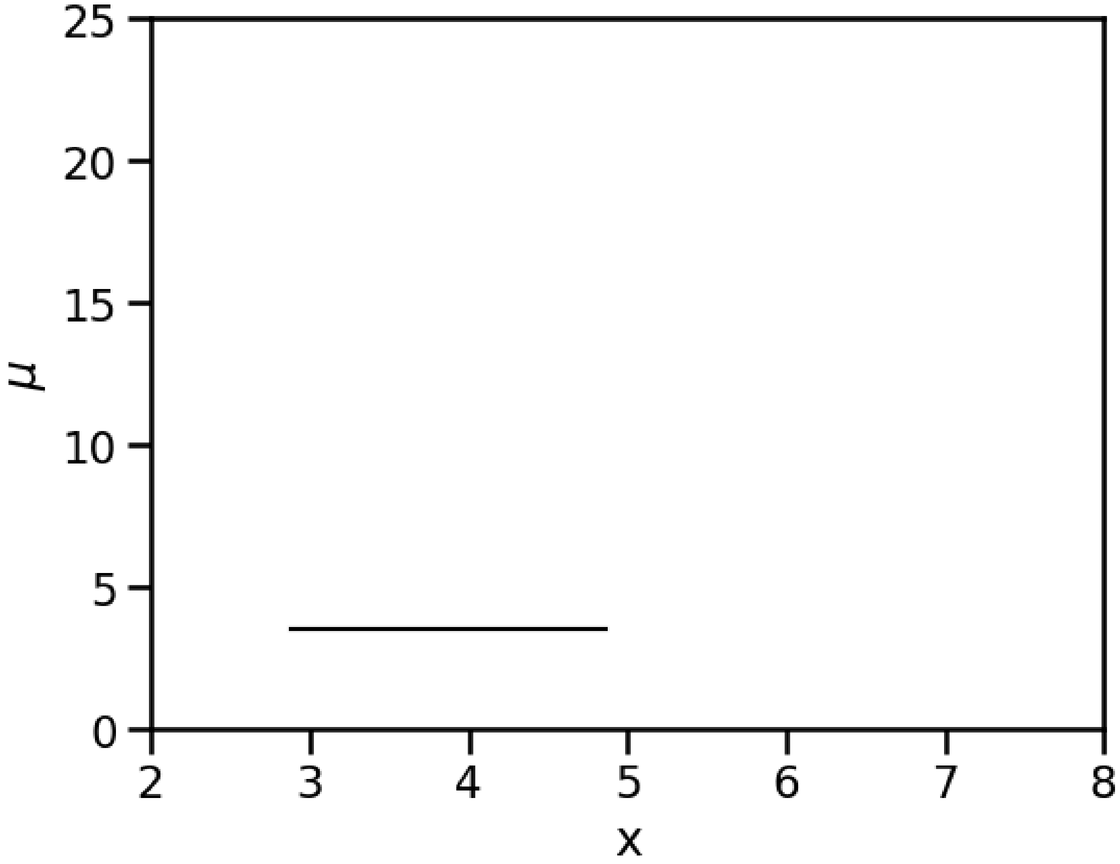
- Recall *likelihood* $\Pr(x|\mu)$ (ie. probability of data, assume known paramater)

CI Construction: Confidence Belt

- Recall *likelihood* $\Pr(x|\mu)$ (ie. probability of data, assume known paramater)
- Take for example $\mu = 4$. Expect 4 neutrinos per time.

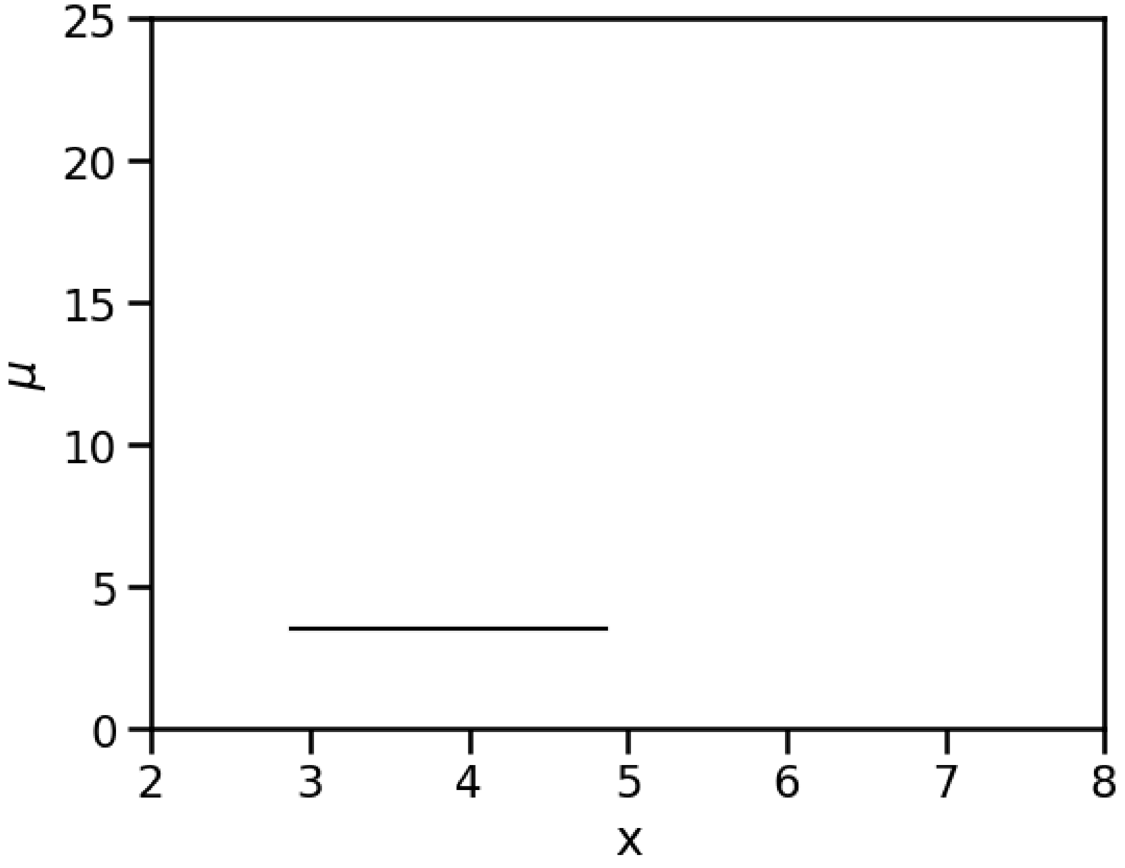
CI Construction: Confidence Belt

- Recall *likelihood* $\Pr(x|\mu)$ (ie. probability of data, assume known paramater)
- Take for example $\mu = 4$. Expect 4 neutrinos per time.



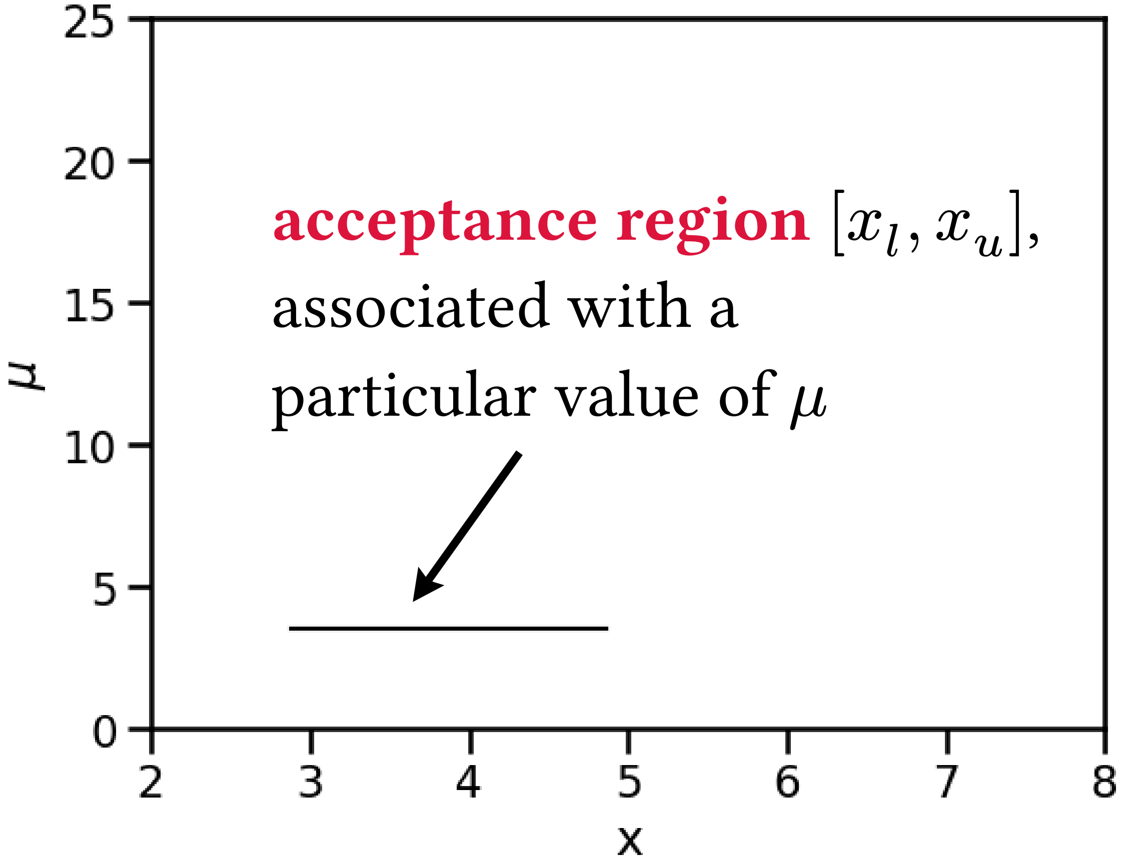
CI Construction: Confidence Belt

- Recall *likelihood* $\Pr(x|\mu)$ (ie. probability of data, assume known paramater)
- Take for example $\mu = 4$. Expect 4 neutrinos per time.
- Select a region $[x_l, x_u]$ such that the probability of measuring $x \in [x_l, x_u]$ is, say, 80%.



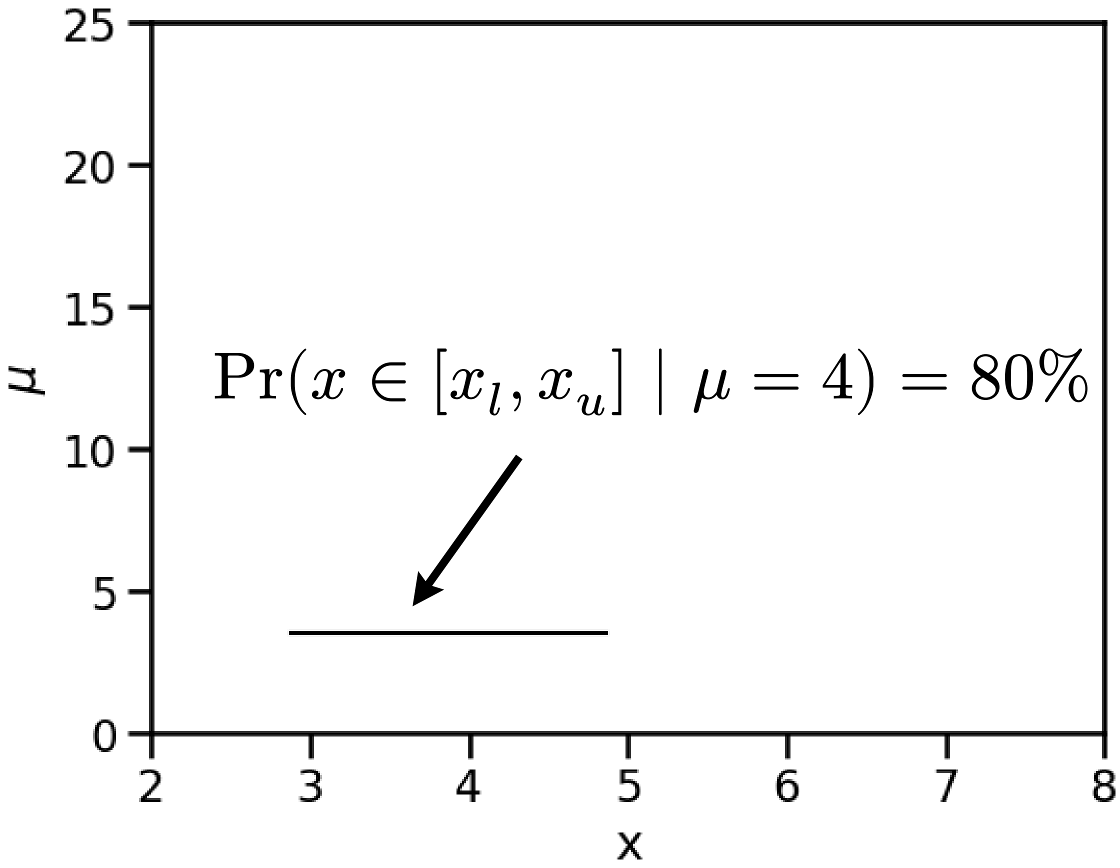
CI Construction: Confidence Belt

- Recall *likelihood* $\Pr(x|\mu)$ (ie. probability of data, assume known paramater)
- Take for example $\mu = 4$. Expect 4 neutrinos per time.
- Select a region $[x_l, x_u]$ such that the probability of measuring $x \in [x_l, x_u]$ is, say, 80%.



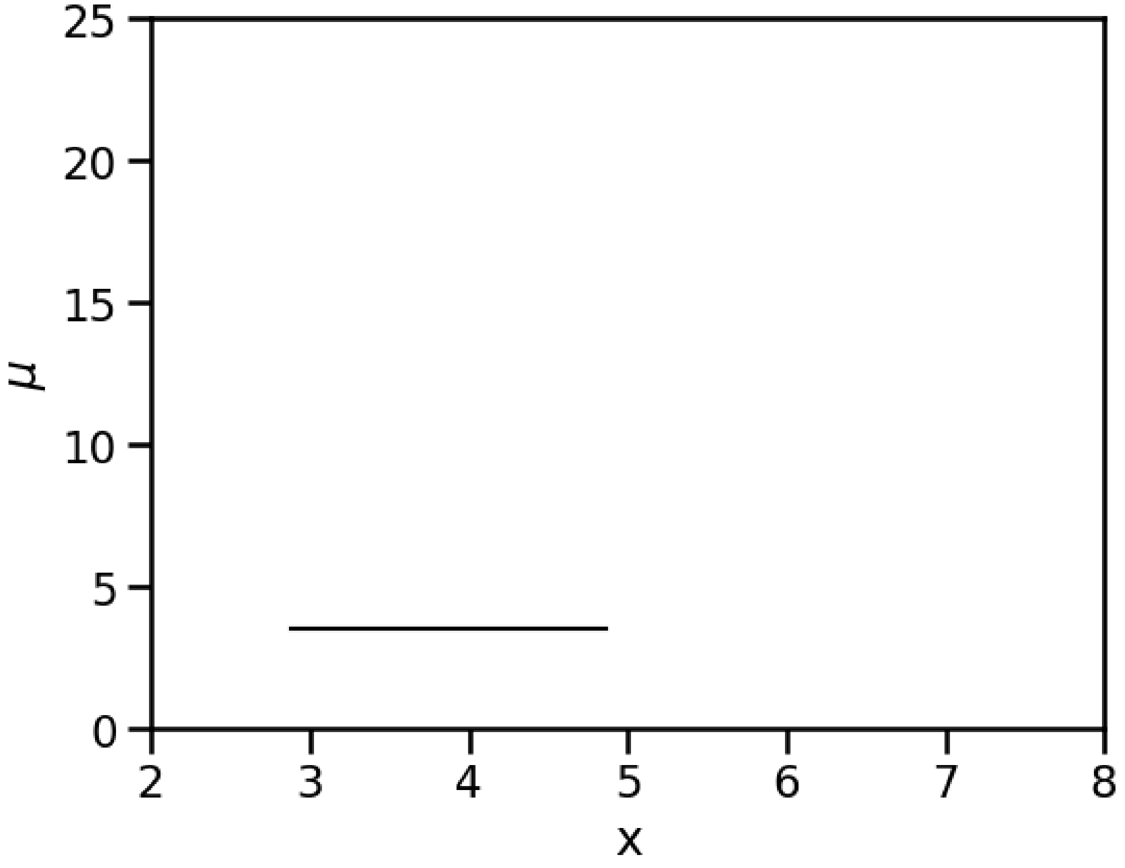
CI Construction: Confidence Belt

- Recall *likelihood* $\Pr(x|\mu)$ (ie. probability of data, assume known paramater)
- Take for example $\mu = 4$. Expect 4 neutrinos per time.
- Select a region $[x_l, x_u]$ such that the probability of measuring $x \in [x_l, x_u]$ is, say, 80%.



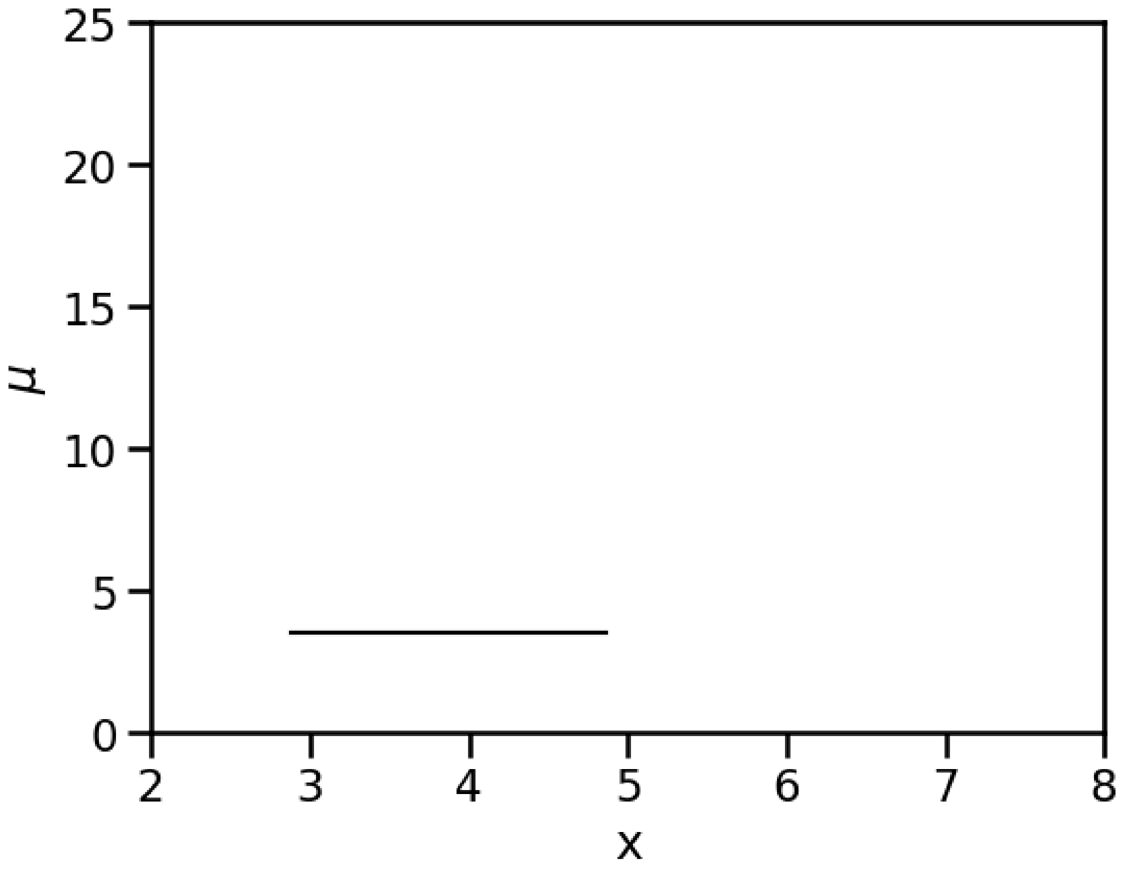
CI Construction: Confidence Belt

- But of course μ_t is unknown.



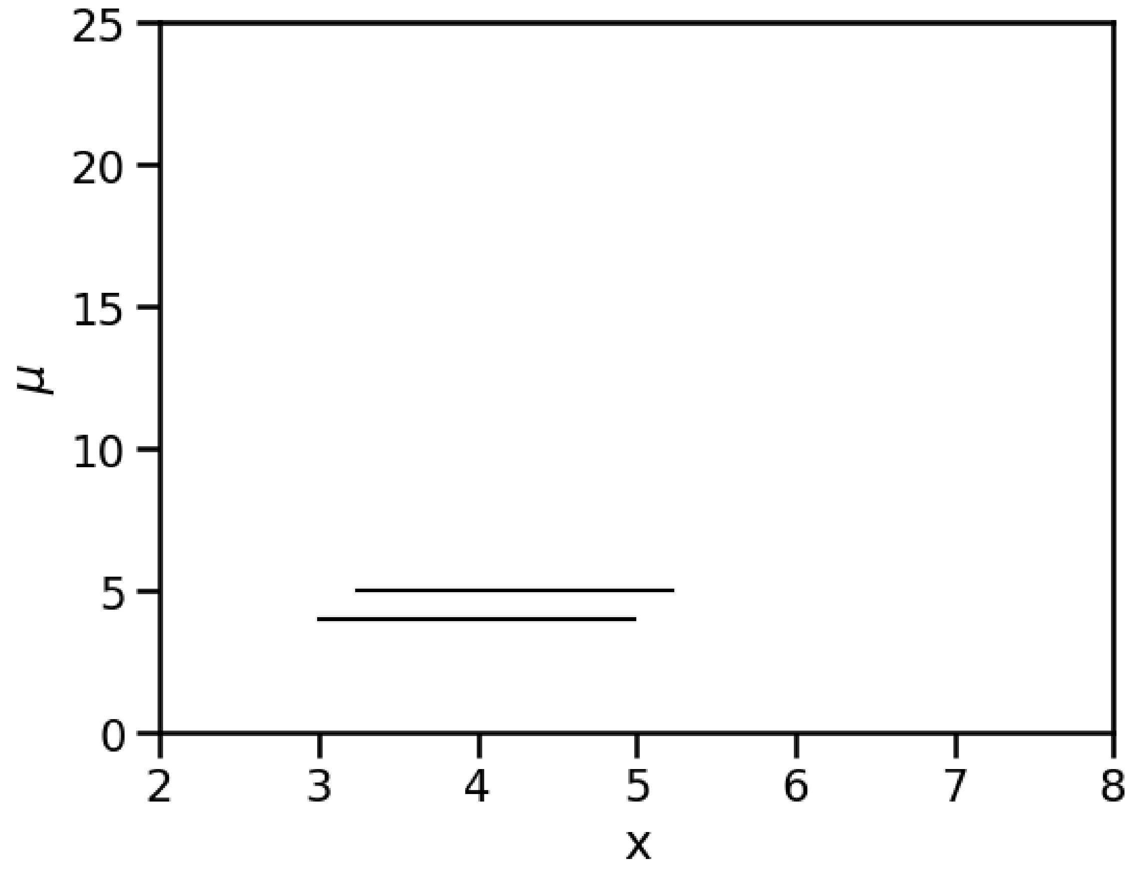
CI Construction: Confidence Belt

- But of course μ_t is unknown.
- The true flux μ_t could be, say, $\mu = 5$



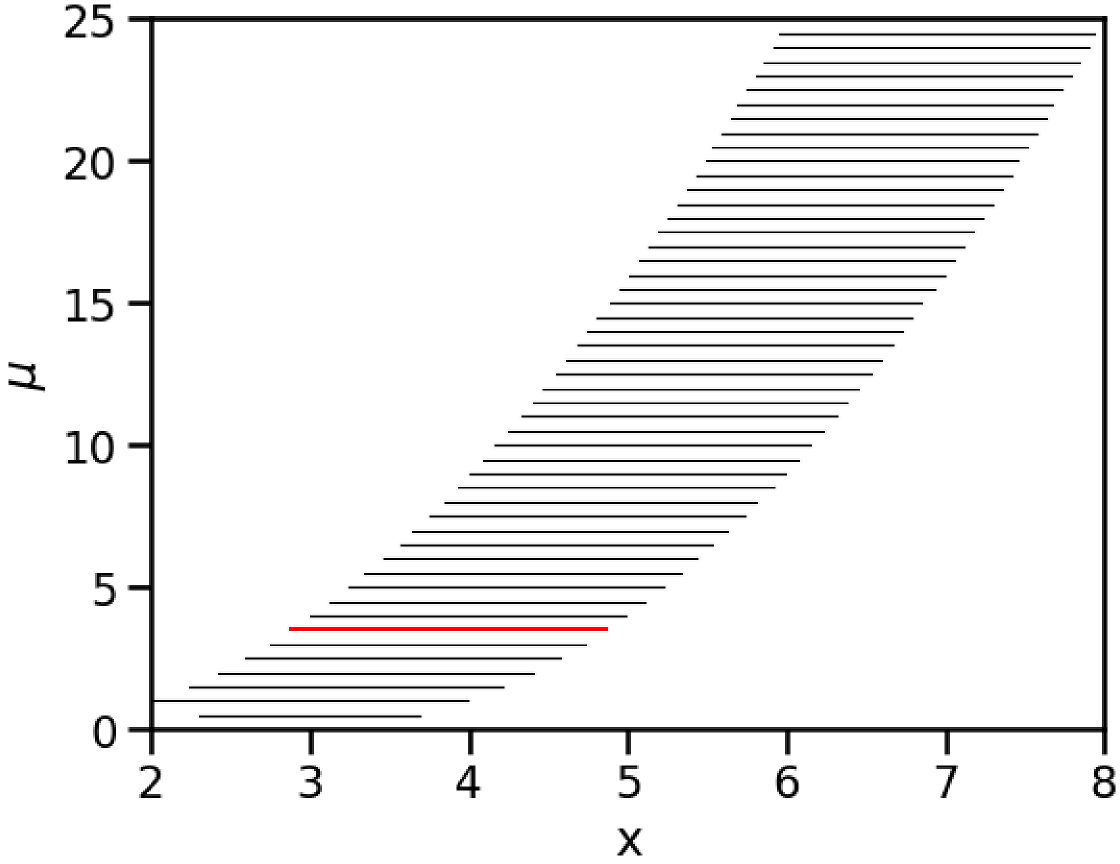
CI Construction: Confidence Belt

- But of course μ_t is unknown.
- The true flux μ_t could be, say, $\mu = 5$
- So, construct another acceptance region $[x_l, x_u]$ for $\mu = 5$



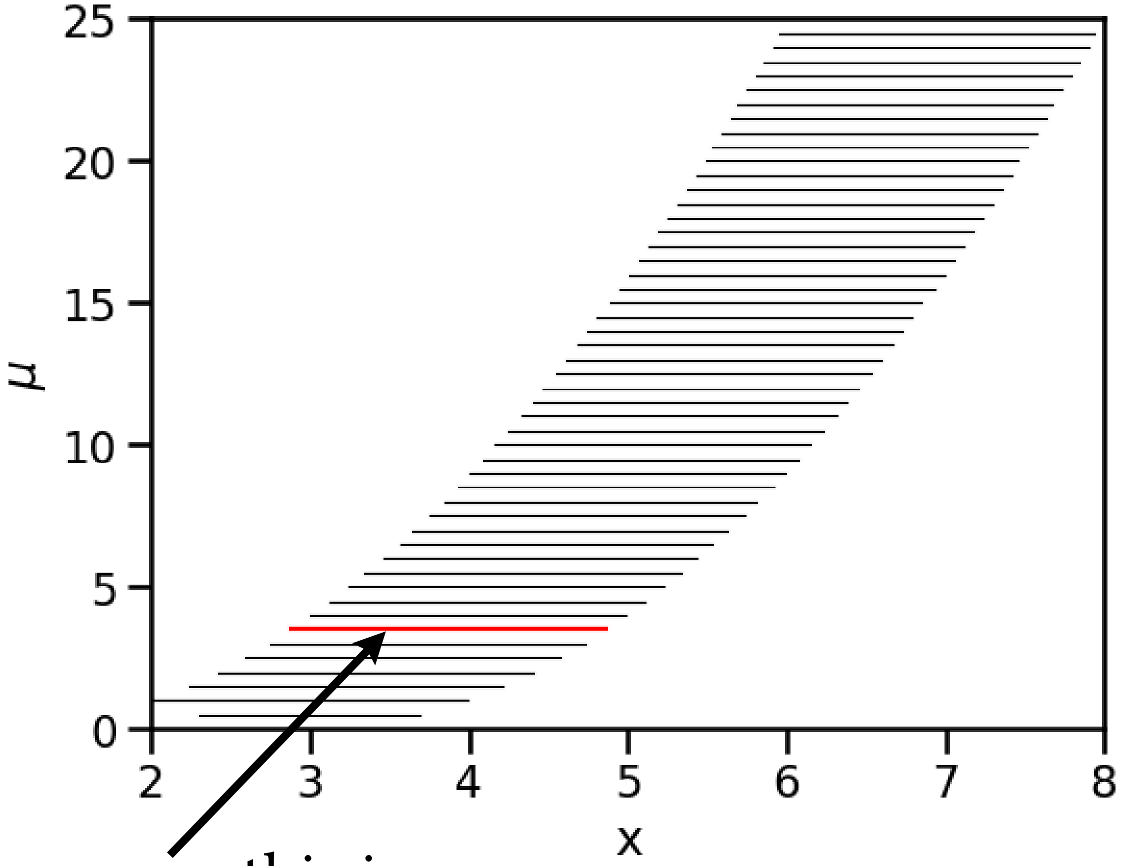
CI Construction: Confidence Belt

- But of course μ_t is unknown.
- The true flux μ_t could be, say, $\mu = 5$
- So, construct another acceptance region $[x_l, x_u]$ for $\mu = 5$
- Rinse and repeat



CI Construction: Confidence Belt

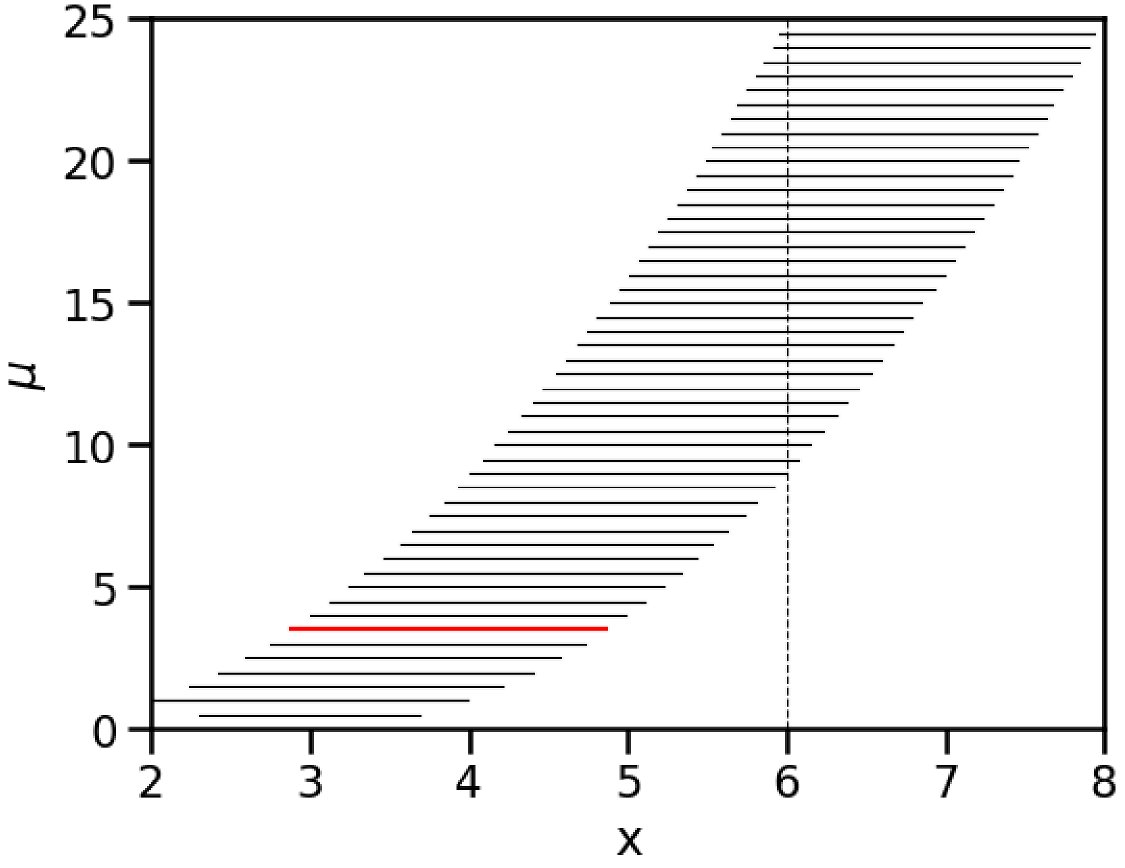
- But of course μ_t is unknown.
- The true flux μ_t could be, say, $\mu = 5$
- So, construct another acceptance region $[x_l, x_u]$ for $\mu = 5$
- Rinse and repeat



suppose this is μ_t

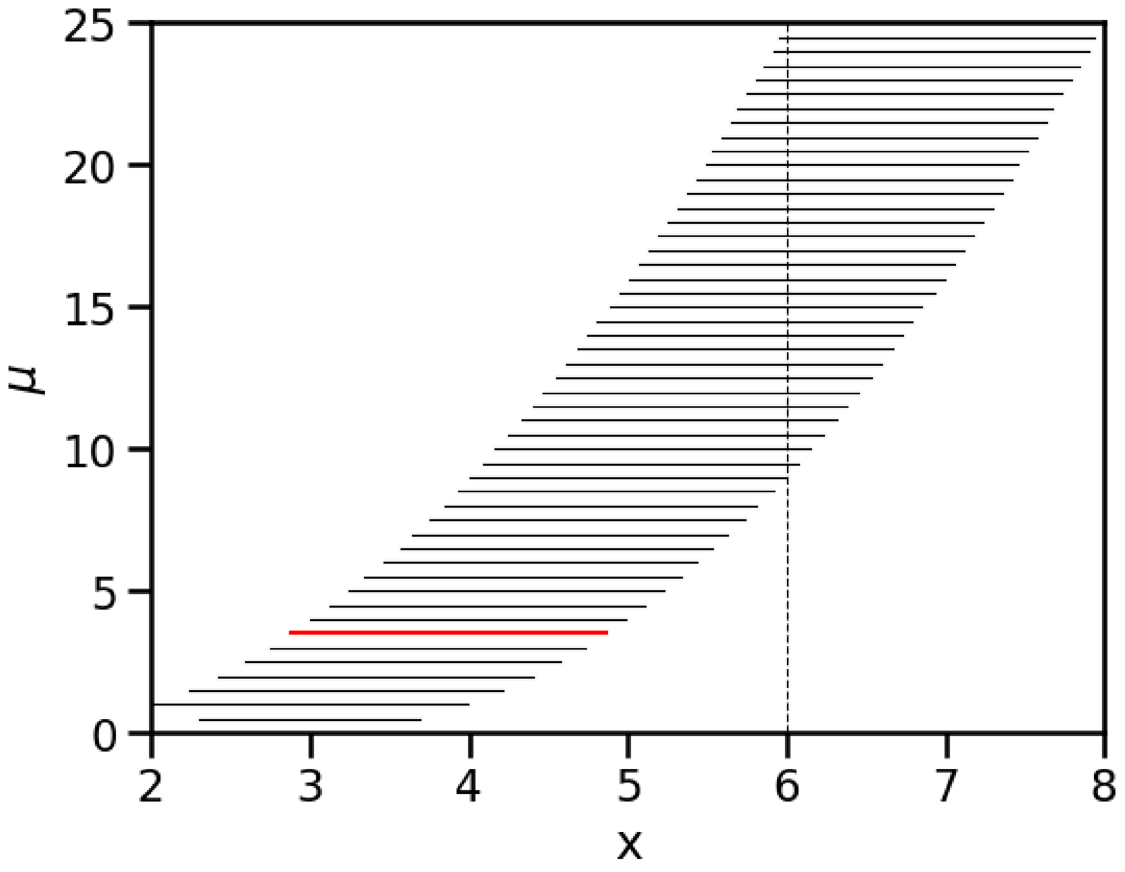
CI Construction: Confidence Belt

- Make a measurement, get result $x_0 = 6$



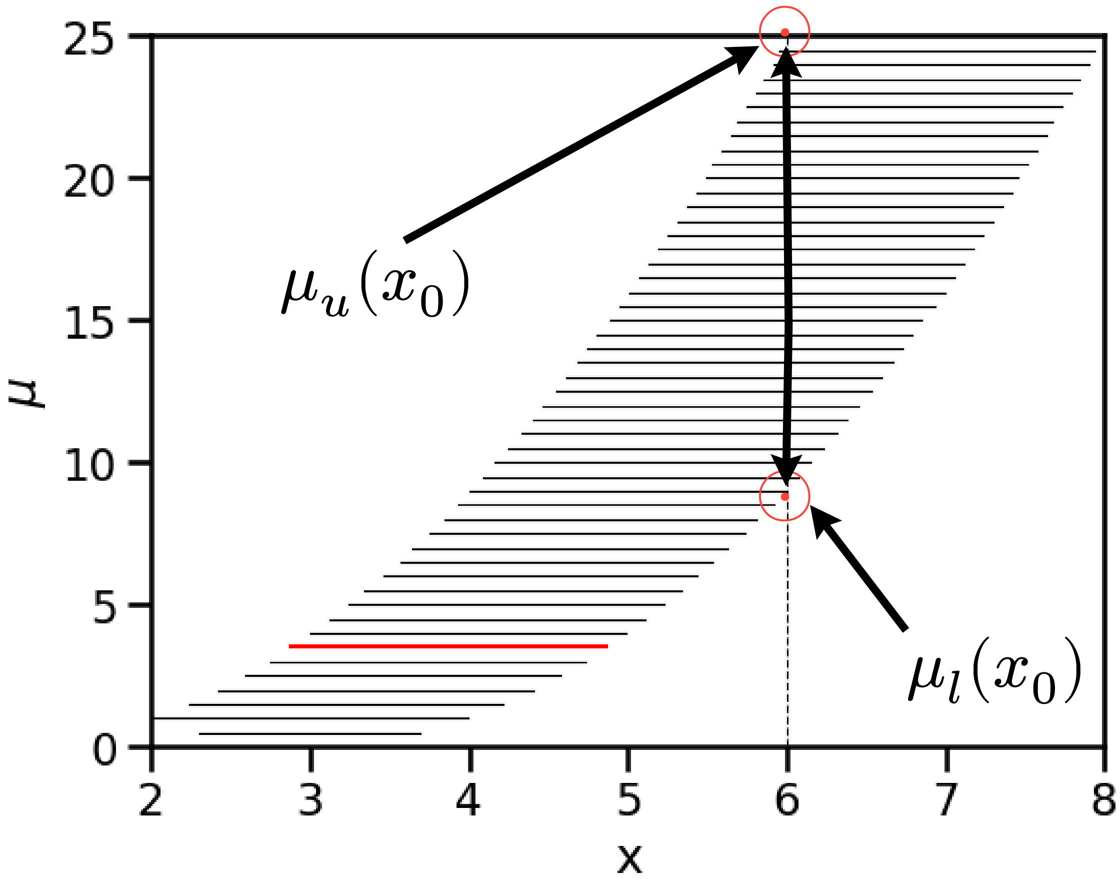
CI Construction: Confidence Belt

- Make a measurement, get result $x_0 = 6$
- The probability of x_0 falling in the acceptance region (red) is 80%, by construction



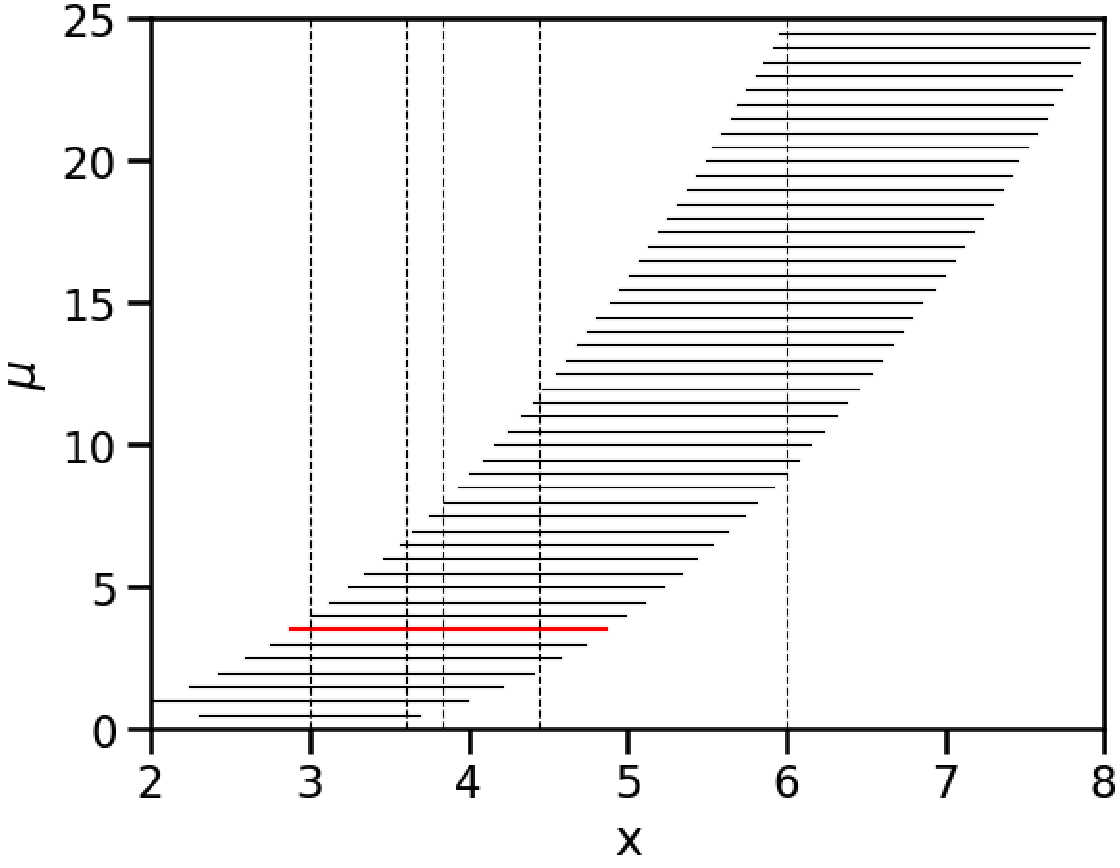
CI Construction: Confidence Belt

- Make a measurement, get result $x_0 = 6$
- The probability of x_0 falling in the acceptance region (red) is 80%, by construction
- The **confidence interval** $[\mu_l, \mu_u]$ from this experiment is the vertical intercept.



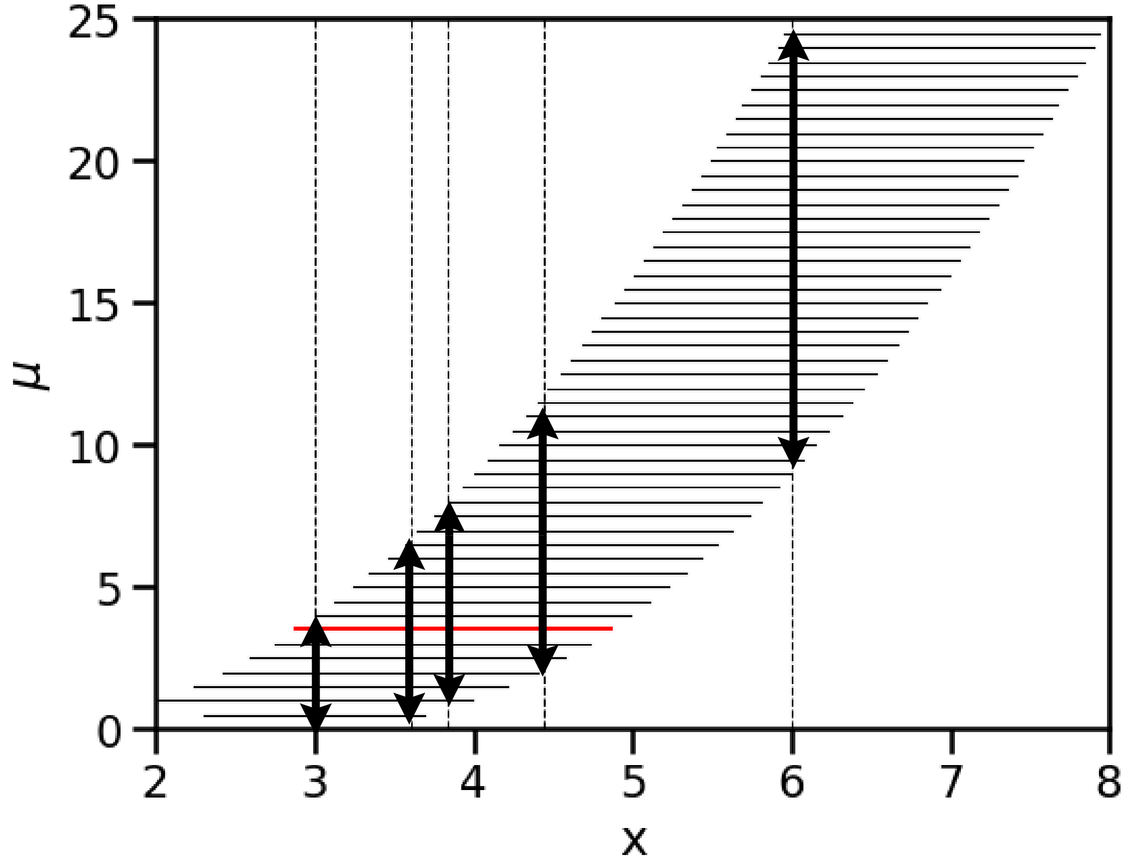
CI Construction: Confidence Belt

- Make some more measurements



CI Construction: Confidence Belt

- Make some more measurements
- Get some more confidence intervals.

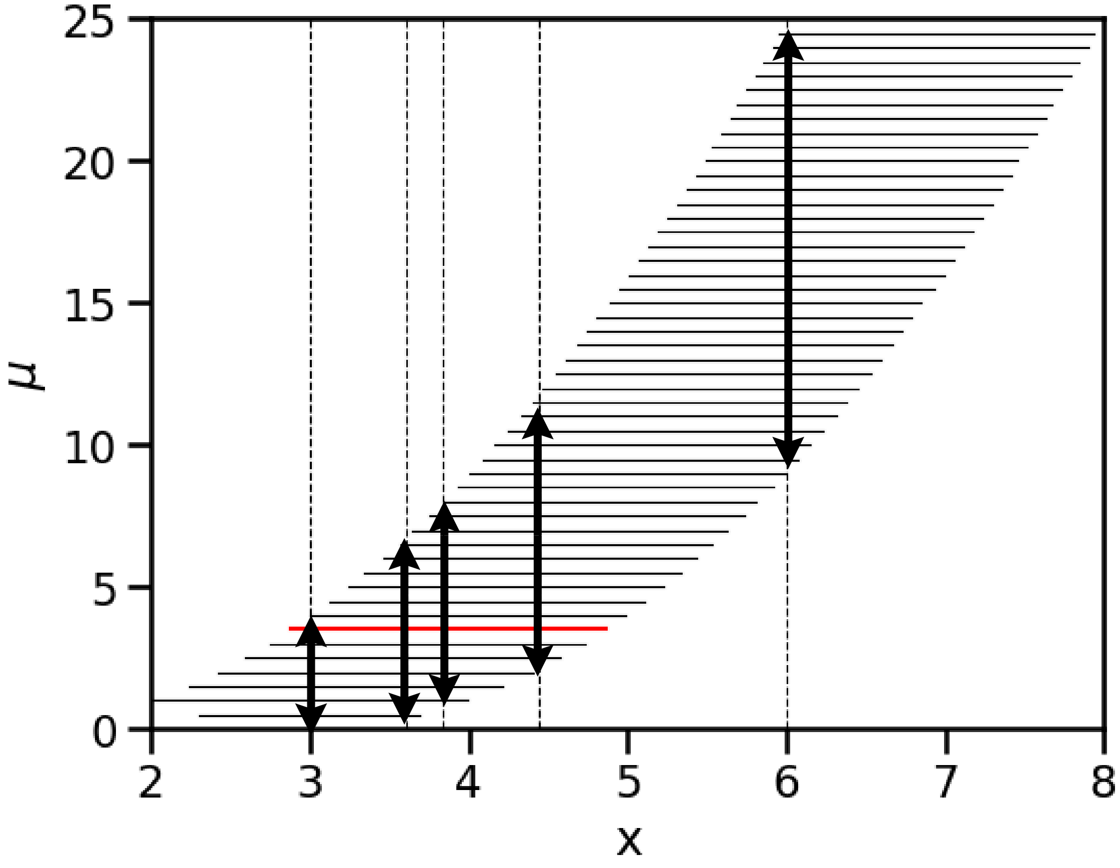


CI Construction: Confidence Belt

- Make some more measurements
- Get some more confidence intervals.
- Have a set

$$C = \{\mathrm{CI}_1, \mathrm{CI}_2, \mathrm{CI}_3, \mathrm{CI}_4, \mathrm{CI}_5\}$$

- 80% of this set would cover the true value, μ_t.



Problem with Standard CI

Standard Poisson (90% Central) CI

- Poisson distribution w/ background b

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

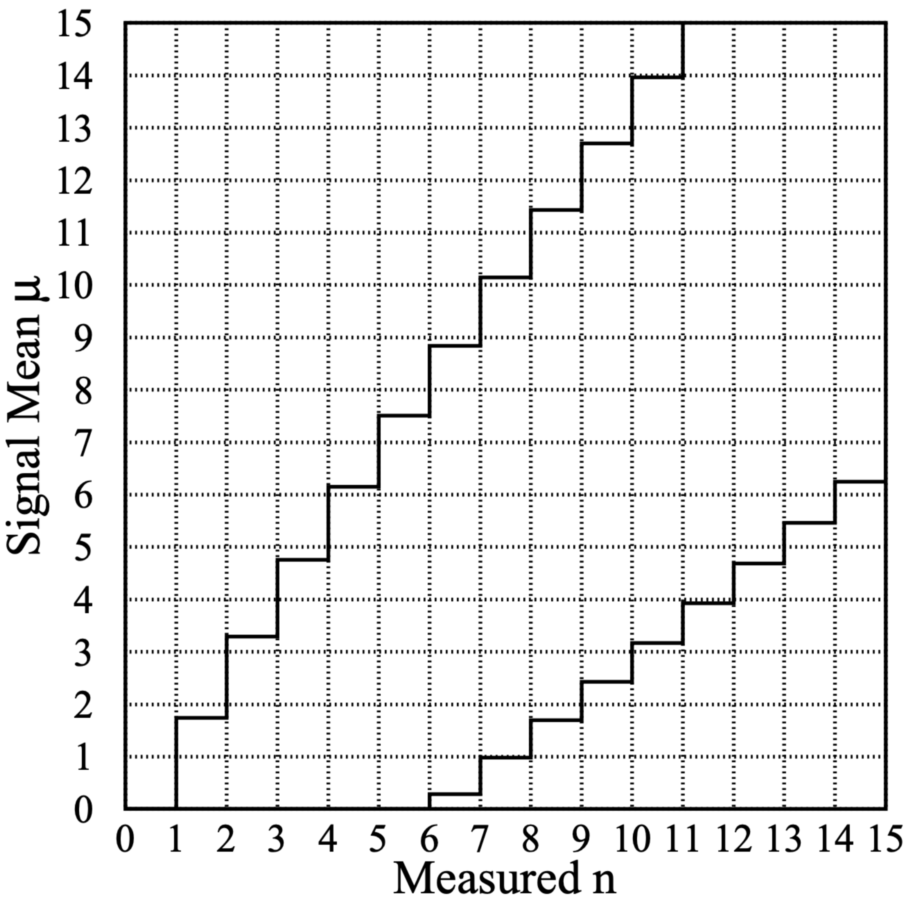
Standard Poisson (90% Central) CI

- Poisson distribution w/ background b

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

- Somehow construct a standard (i.e. "classical") 90% **central** interval
  - ref. Feldman & Cousins Section II.B and Fig. 6

Standard Poisson (90% Central) CI

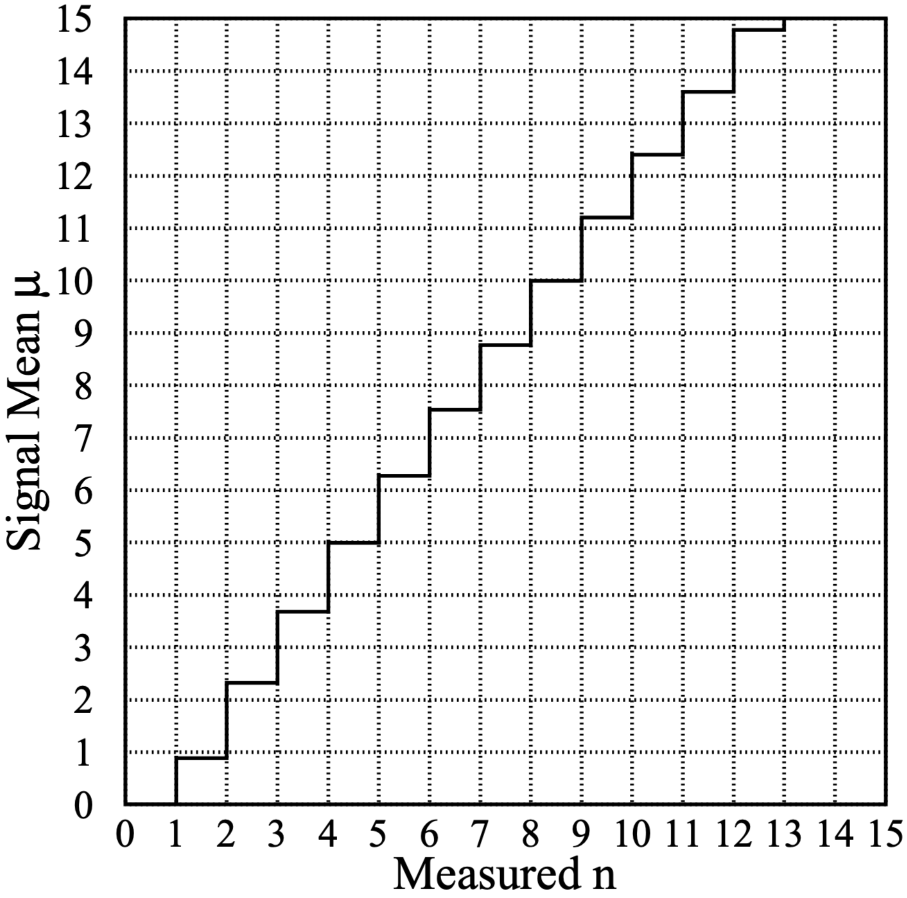


- Poisson distribution w/ background b

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

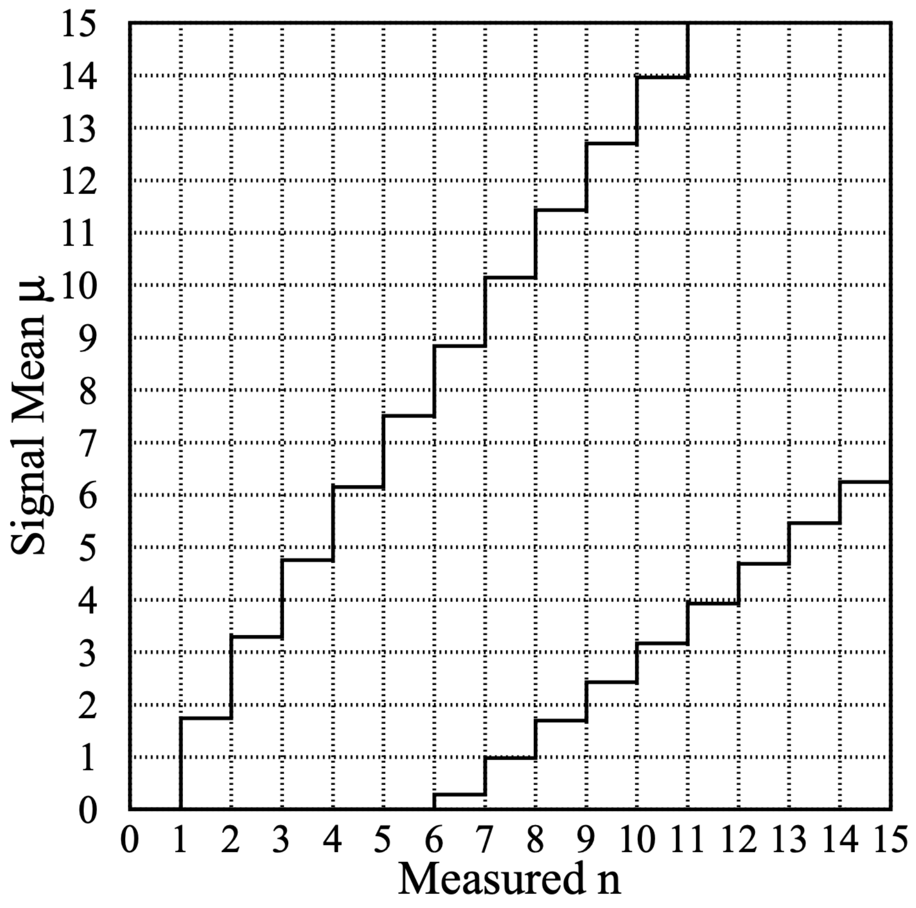
- Somehow construct a standard (i.e. "classical") 90% **central** interval
  - ref. Feldman & Cousins Section II.B and Fig. 6

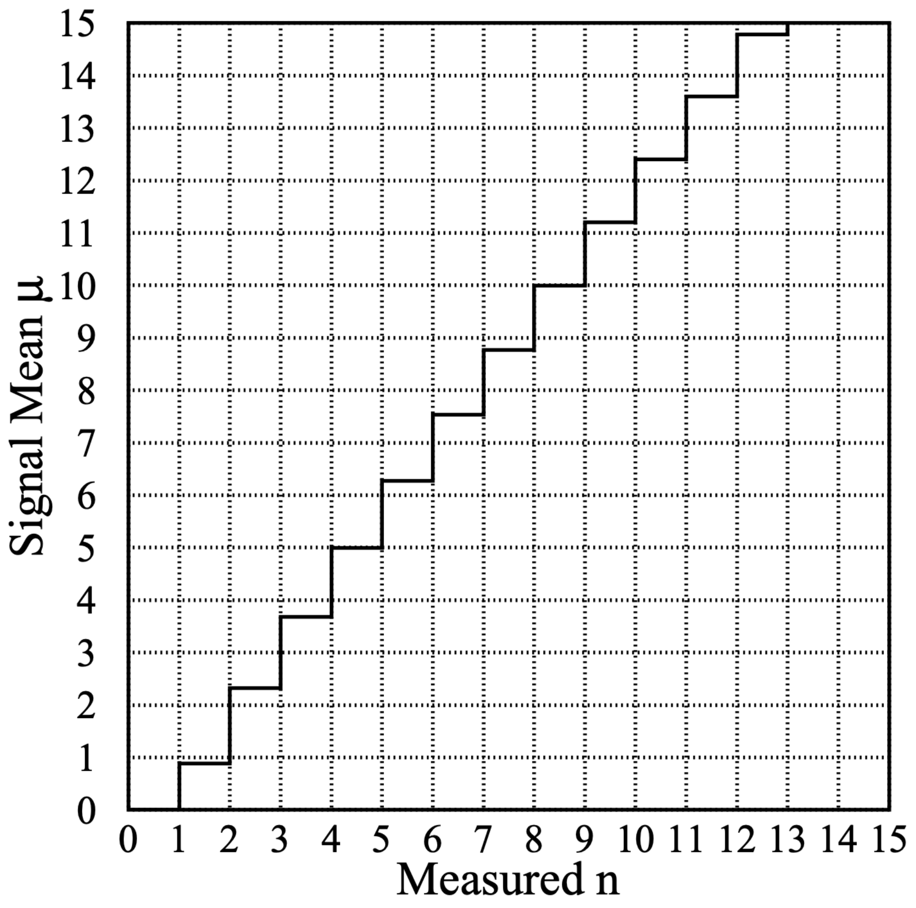
Standard Poisson (90% Upper) CI



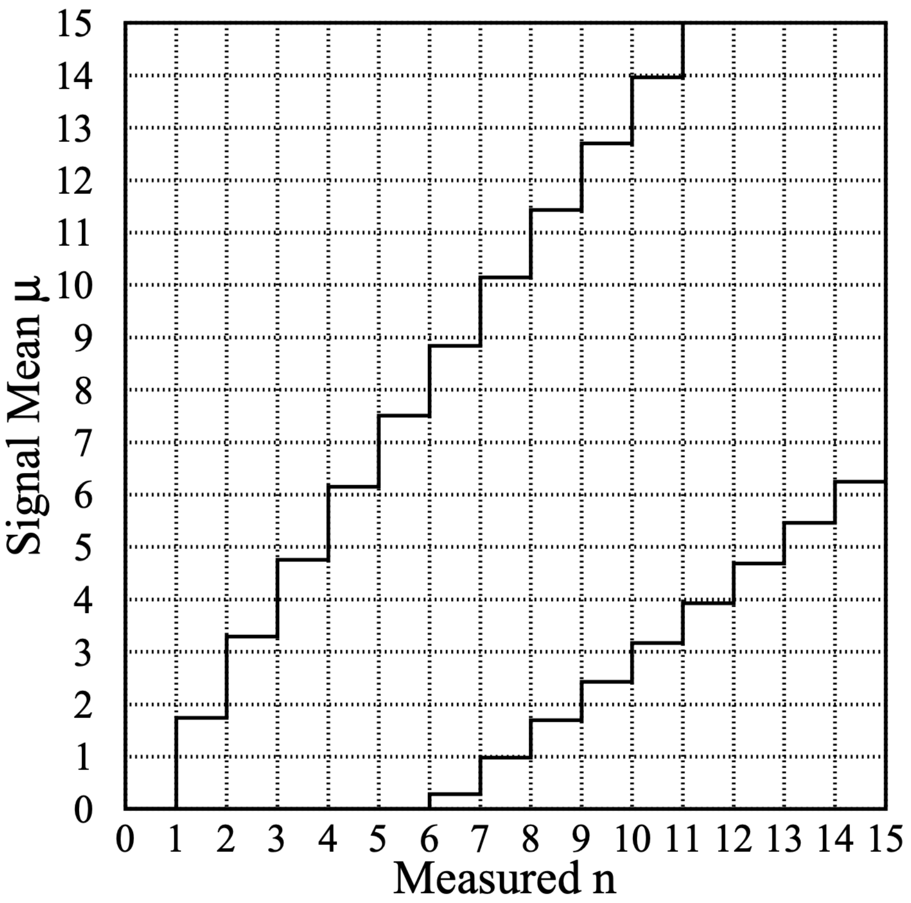
- Similarly, a standard 90% **upper** limit can be constructed.
- ref. Feldman & Cousins Fig. 5

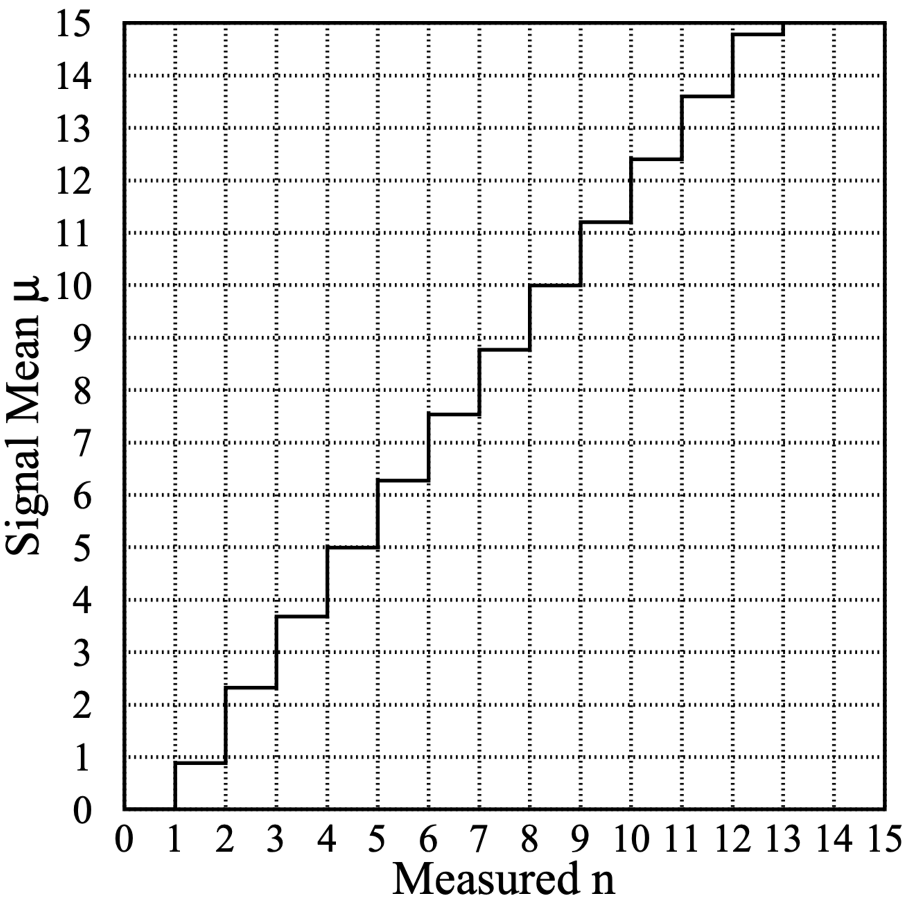
Problem: “flip-flop”





Problem: “flip-flop”



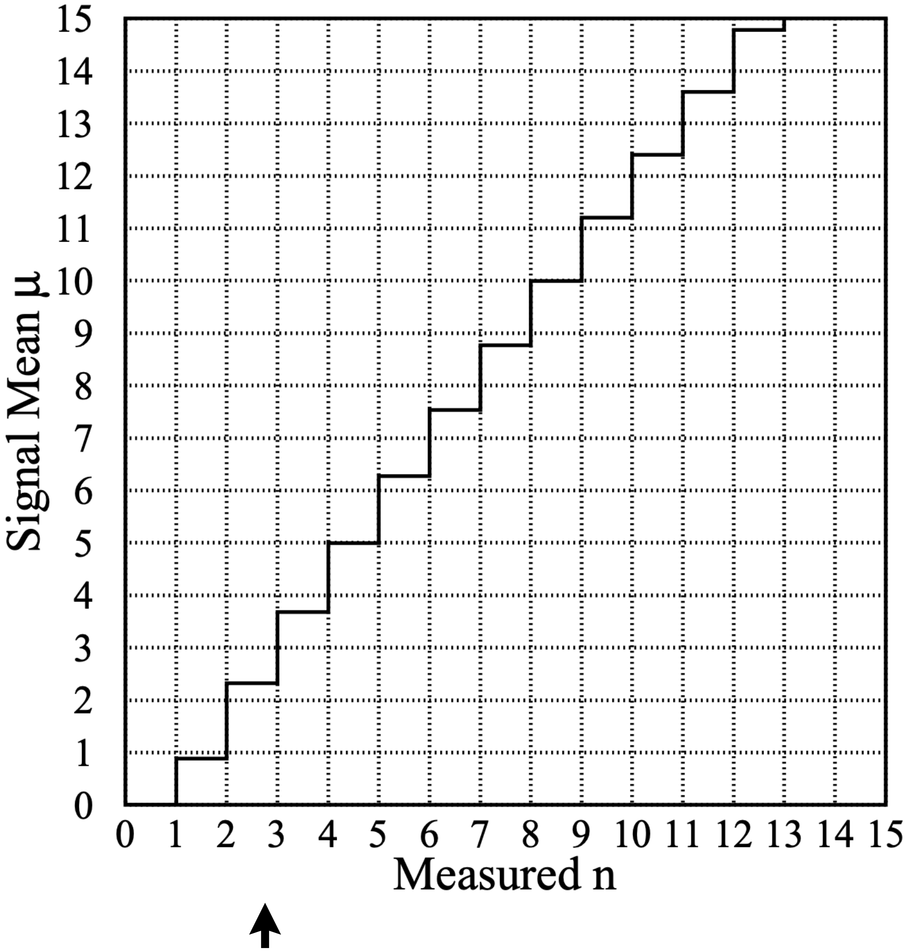


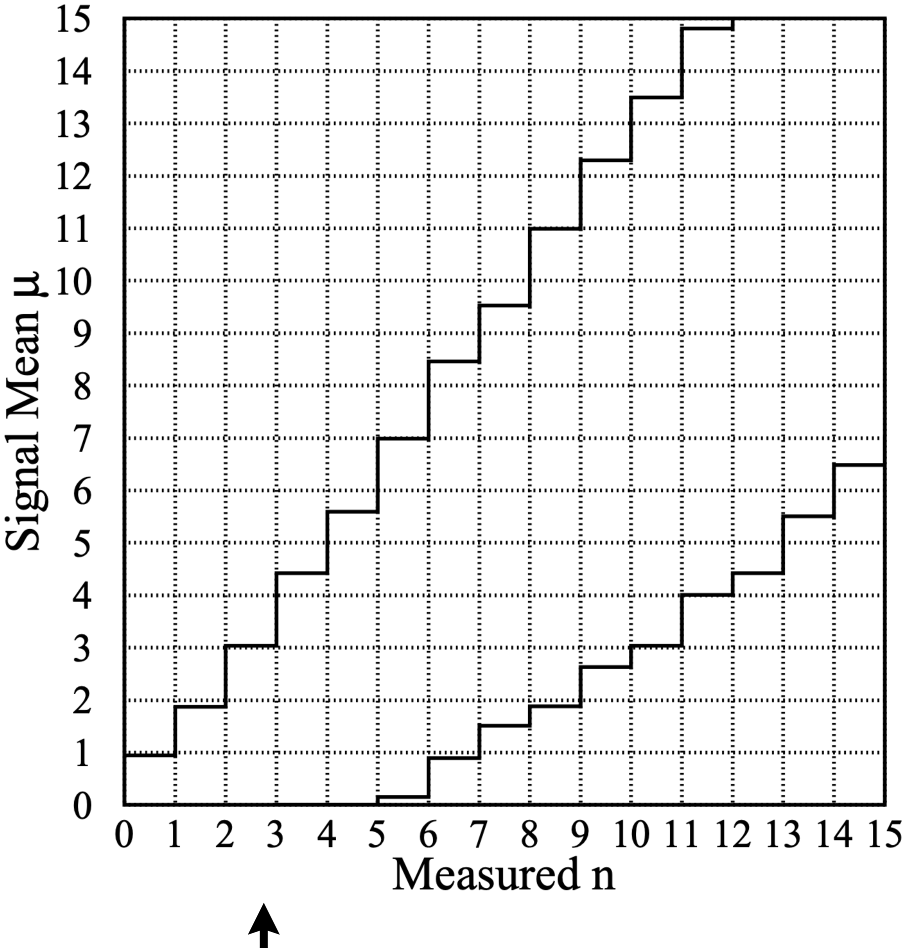
One **cannot** pick which interval to report after-the-fact.

Problem: “flip-flop”

- It turns out that flip-flopping leads to invalid intervals
- Feldman & Cousins’s approach removes the possibility of (or motivation to commit) flip-flopping.

Comparison: look at small n

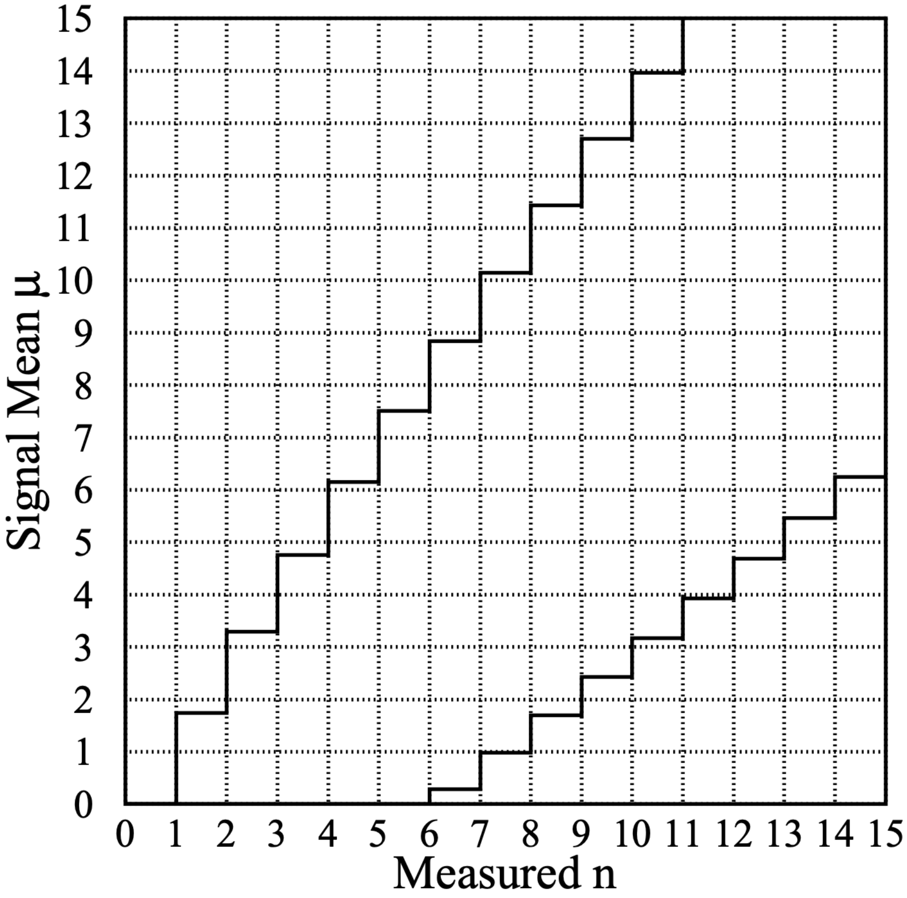


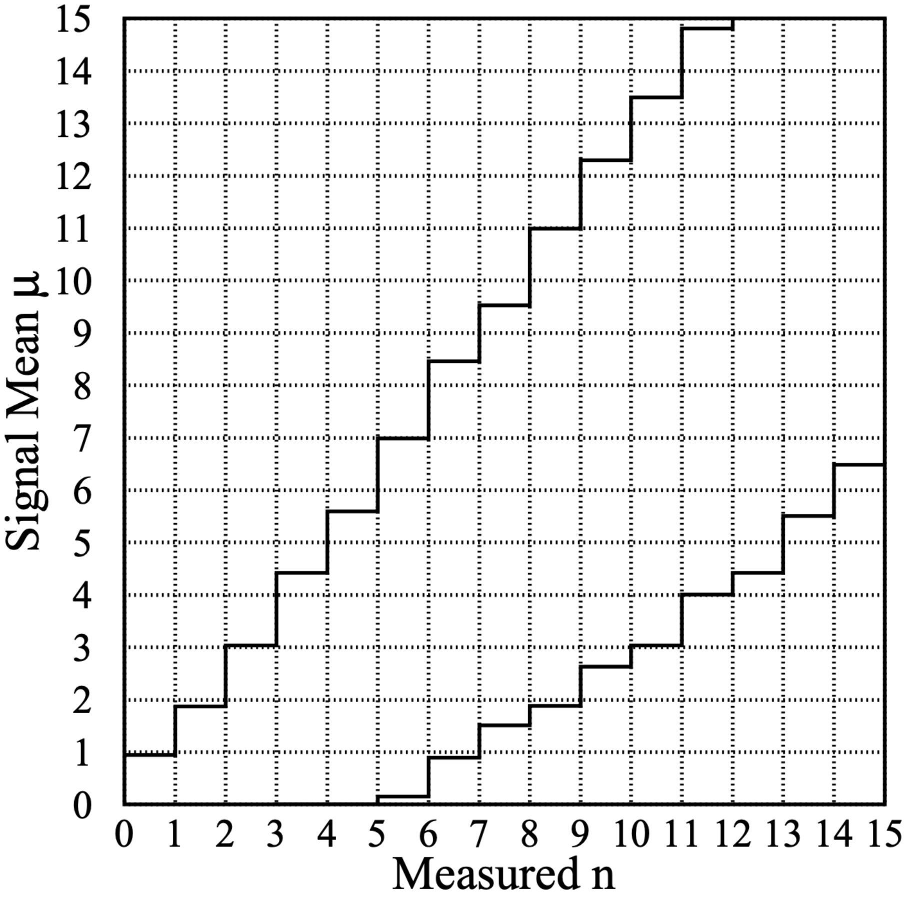


standard 90% upper CI

Feldman & Cousins

Comparison: look at large n



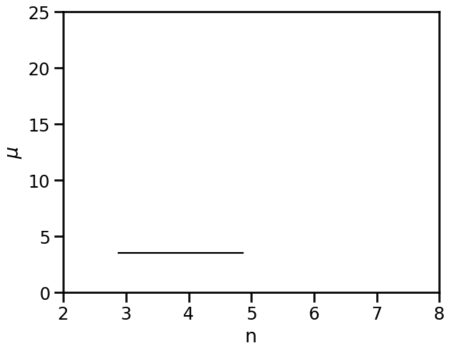


standard 90% central CI

Feldman & Cousins

Acceptance Region

Maximum Likelihood



- Recall acceptance region:

$$\Pr(n \in [n_1, n_2] \mid \mu_{\text{fixed}}) = 90\%$$

- Complete freedom in choosing how to construct the acceptance regions.
- Consider likelihood: Poisson with background b:

$$\mathcal{L} \equiv \Pr(n \mid \mu) = \frac{(\mu + b)^n e^{-(\mu+b)}}{n!}$$

- F&C propose to compute a likelihood ratio R
  - This needs a "best fit" $\mu_{\text{best}} \equiv \max(0, n - b)$

Maximum Likelihood

Derivation (skip me!)

- Likelihood is a Poisson in this case.

$$\mathcal{L} \equiv \Pr(n \mid \mu) = \frac{(\mu + b)^n e^{-(\mu+b)}}{n!}$$

- Find maximum (fixing n, vary μ):

$$\left.\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\mu}\right|_{\mu=\mu_{\text{best}}} = 0$$

- Result: “best fit” $\mu = \mu_{\text{best}} = n - b$
- Require physical $\mu \geq 0 \Rightarrow \mu_{\text{best}} = \max(0, n - b)$

Likelihood Ratio

- Do this for representative values of μ; say we start with $\mu = 0.5$
  - As an example background, $b = 3$
  - $\Rightarrow \mu_{\text{best}} \equiv \max(0, n - b) = \max(0, n - 3)$
- Procedure:
  -
  -
  -
  -

n	$\Pr(n\|\mu)$	μ_{best}	$\Pr(n\|\mu_{\text{best}})$	R	rank
0					

Likelihood Ratio

- Do this for representative values of μ; say we start with $\mu = 0.5$
  - As an example background, $b = 3$
  - $\Rightarrow \mu_{\text{best}} \equiv \max(0, n - b) = \max(0, n - 3)$
- Procedure:
  - For $n = 0$, compute $\Pr(n \mid \mu = 0.5)$
  -
  -
  -

n	$\Pr(n\|\mu)$	μ_{best}	$\Pr(n\|\mu_{\text{best}})$	R	**rank**
0	0.03				

Likelihood Ratio

- Do this for representative values of μ; say we start with $\mu = 0.5$
  - As an example background, $b = 3$
  - $\Rightarrow \mu_{\text{best}} \equiv \max(0, n - b) = \max(0, n - 3)$
- Procedure:
  - For $n = 0$, compute $\Pr(n \mid \mu = 0.5)$
  - For $n = 0$, compute $\mu_{\text{best}} = 0$
  -
  -

n	$\Pr(n\|\mu)$	μ_{best}	$\Pr(n\|\mu_{\text{best}})$	R	rank
0	0.03	0			

Likelihood Ratio

- Do this for representative values of μ; say we start with $\mu = 0.5$
  - As an example background, $b = 3$
  - $\Rightarrow \mu_{\text{best}} \equiv \max(0, n - b) = \max(0, n - 3)$
- Procedure:
  - For $n = 0$, compute $\Pr(n \mid \mu = 0.5)$
  - For $n = 0$, compute $\mu_{\text{best}} = 0$
  - For $n = 0$, compute $\Pr(n \mid \mu = \mu_{\text{best}})$
  -

n	$\Pr(n\|\mu)$	μ_{best}	$\Pr(n\|\mu_{\text{best}})$	R	rank
0	0.03	0	0.05		

Likelihood Ratio

- Do this for representative values of μ; say we start with $\mu = 0.5$
  - As an example background, $b = 3$
  - $\Rightarrow \mu_{\text{best}} \equiv \max(0, n - b) = \max(0, n - 3)$
- Procedure:
  - For $n = 0$, compute $\Pr(n \mid \mu = 0.5)$
  - For $n = 0$, compute $\mu_{\text{best}} = 0$
  - For $n = 0$, compute $\Pr(n \mid \mu = \mu_{\text{best}})$
  - Divide likelihoods to get R.

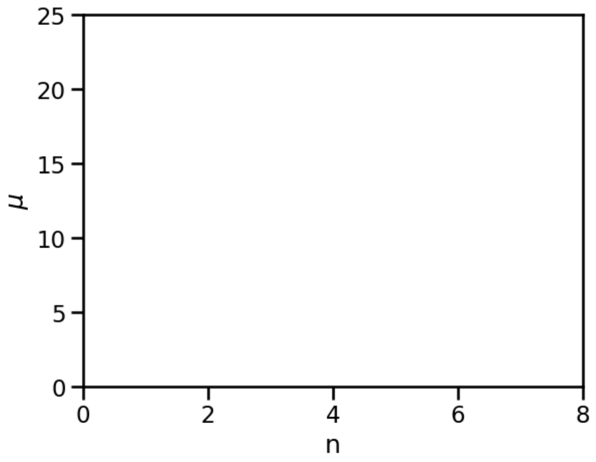
n	$\Pr(n\|\mu)$	μ_{best}	$\Pr(n\|\mu_{\text{best}})$	R	rank
0	0.03	0	0.05	0.607	

Likelihood Ratio

- As a reminder, this is still just for $\mu = 0.5$ (and example $b = 3$)

n	$\Pr(n\|\mu)$	μ_{best}	$\Pr(n\|\mu_{\text{best}})$	R	rank
0	0.03	0	0.05	0.607	

- We will see that the acceptance region for $\mu = 0.5$ is this:

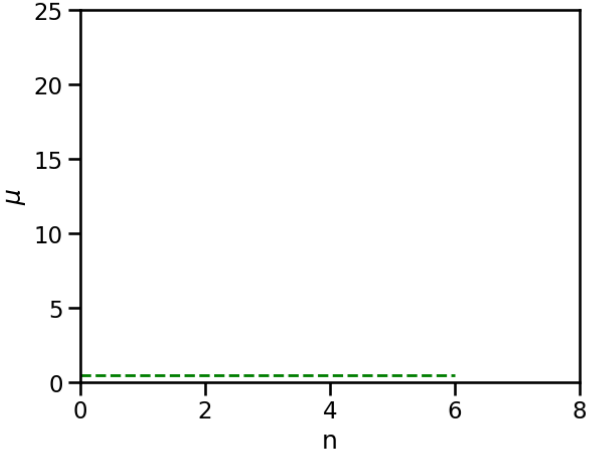


Likelihood Ratio

- As a reminder, this is still just for $\mu = 0.5$ (and example $b = 3$)

n	$\Pr(n\|\mu)$	μ_{best}	$\Pr(n\|\mu_{\text{best}})$	R	rank
0	0.03	0	0.05	0.607	

- We will see that the acceptance region for $\mu = 0.5$ is this:



Likelihood Ratio

- As a reminder, this is still just for $\mu = 0.5$ (and example $b = 3$)

n	$\Pr(n\|\mu)$	μ_{best}	$\Pr(n\|\mu_{\text{best}})$	R	rank
0	0.03	0	0.05	0.607	

- To construct the region, make a new row for $n = 1$

n	$\Pr(n\|\mu)$	μ_{best}	$\Pr(n\|\mu_{\text{best}})$	R	rank
0	0.030	0	0.050	0.607	
1	0.106	0	0.149	0.708	

Example acceptance region for $\mu = 0.5$

And then for a bunch of other n.

Example acceptance region for $\mu = 0.5$

n	$\Pr(n\|\mu)$	μ_{best}	$\Pr(n\|\mu_{\text{best}})$	R	rank
0	0.030	0	0.050	0.607	
1	0.106	0	0.149	0.708	
2	0.185	0	0.224	0.826	
3	0.216	0	0.224	0.963	
4	0.189	1	0.195	0.966	
5	0.132	2	0.175	0.753	
6	0.077	3	0.161	0.480	
7	0.039	4	0.149	0.259	

Example acceptance region for $\mu = 0.5$

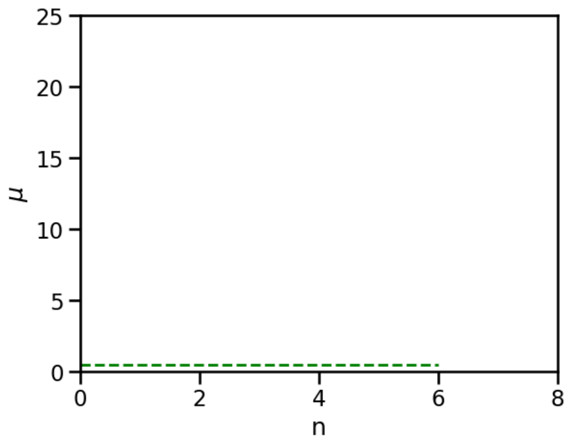
n	$\Pr(n\|\mu)$	μ_{best}	$\Pr(n\|\mu_{\text{best}})$	R	rank
0	0.030	0	0.050	0.607	6
1	0.106	0	0.149	0.708	5
2	0.185	0	0.224	0.826	3
3	0.216	0	0.224	0.963	2
4	0.189	1	0.195	0.966	1
5	0.132	2	0.175	0.753	4
6	0.077	3	0.161	0.480	7
7	0.039	4	0.149	0.259	

Example acceptance region for $\mu = 0.5$

- Start adding $\Pr(n|\mu)$ in the second column based on the rank.
- Stop when total probability exceeds 90%.
- The n's that contribute to the sum are the ones included in the acceptance region.

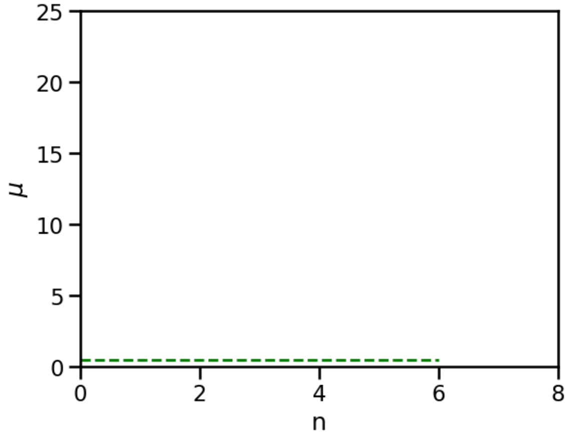
Example acceptance region for $\mu = 0.5$

- Start adding $\Pr(n|\mu)$ in the second column based on the rank.
- Stop when total probability exceeds 90%.
- The n's that contribute to the sum are the ones included in the acceptance region.
- Acceptance region for $\mu = 0.5$ is therefore $n \in [0, 6]$



Example acceptance region for $\mu = 0.5$

- Start adding $\Pr(n|\mu)$ in the second column based on the rank.
- Stop when total probability exceeds 90%.
- The n's that contribute to the sum are the ones included in the acceptance region.
- Acceptance region for $\mu = 0.5$ is therefore $n \in [0, 6]$
- Next, construct the acceptance region for other μ as well.



Bayesian

Bayes Theorem

- The problem is that the parameter λ is what we want to figure out.

Bayes Theorem

- The problem is that the parameter λ is what we want to figure out.
  - *e.g.* the flux of neutrinos at some energy $E = 10^{21}$ eV.

Bayes Theorem

- The problem is that the parameter λ is what we want to figure out.
  - *e.g.* the flux of neutrinos at some energy $E = 10^{21}$ eV.
- What we can do is measure n and *estimate* λ

Bayes Theorem

- The problem is that the parameter λ is what we want to figure out.
  - *e.g.* the flux of neutrinos at some energy $E = 10^{21}$ eV.
- What we can do is measure n and *estimate* λ
- Bayes Theorem

$$\boxed{\Pr(\lambda \mid n)} = \frac{\Pr(n \mid \lambda) \cdot \Pr(\lambda)}{\Pr(n)}$$

Posterior

Bayes Theorem

- The problem is that the parameter λ is what we want to figure out.
  - *e.g.* the flux of neutrinos at some energy $E = 10^{21}$ eV.
- What we can do is measure n and *estimate* λ
- Bayes Theorem

Likelihood $\frac{e^{-\lambda}\lambda^n}{n!}$

$$\Pr(\lambda \mid n) = \frac{\Pr(n \mid \lambda) \cdot \Pr(\lambda)}{\Pr(n)}$$

Posterior

Bayes Theorem

- The problem is that the parameter λ is what we want to figure out.
  - *e.g.* the flux of neutrinos at some energy $E = 10^{21}$ eV.
- What we can do is measure n and *estimate* λ
- Bayes Theorem

Likelihood $\frac{e^{-\lambda}\lambda^n}{n!}$

$$\Pr(\lambda \mid n) = \frac{\Pr(n \mid \lambda) \cdot \Pr(\lambda)}{\Pr(n)}$$

Posterior

evidence

- Evidence is typically just a normalization and ignored. Let's call it 1 :)

Bayes Theorem

$$\Pr(\lambda \mid n) = \frac{\Pr(n \mid \lambda) \cdot \Pr(\lambda)}{\Pr(n)}$$

Bayes Theorem

$$\Pr(\lambda \mid n) = \frac{\Pr(n \mid \lambda) \cdot \boxed{\Pr(\lambda)}}{\Pr(n)} \longleftarrow \textit{prior}$$

- If we specify a prior, we get the posterior.

Bayes Theorem

$$\Pr(\lambda \mid n) = \frac{\Pr(n \mid \lambda) \cdot \Pr(\lambda)}{\Pr(n)} \longleftarrow \textit{prior}$$

- If we specify a prior, we get the posterior.
- "uniform prior" $\Pr(\lambda) = 1$

Bayes Theorem

$$\Pr(\lambda \mid n) = \frac{\Pr(n \mid \lambda) \cdot \boxed{\Pr(\lambda)}}{\Pr(n)} \longleftarrow \textit{prior}$$

- If we specify a prior, we get the posterior.
- "uniform prior" $\Pr(\lambda) = 1 \Longrightarrow \Pr(\lambda \mid n) = \Pr(n \mid \lambda)$

Bayes Theorem

$$\Pr(\lambda \mid n) = \frac{\Pr(n \mid \lambda) \cdot \boxed{\Pr(\lambda)}}{\Pr(n)} \longleftarrow \textit{prior}$$

- If we specify a prior, we get the posterior.
- "uniform prior" $\Pr(\lambda) = 1 \Longrightarrow \Pr(\lambda \mid n) = \Pr(n \mid \lambda)$
- Suppose we measure $n = 0$ event, then the posterior is

$$\Pr(\lambda \mid n = 0) = \Pr(n = 0|\lambda) = \frac{e^{-\lambda} \cdot \lambda^n}{n!} = e^{-\lambda}$$

Summary: Bayesian Credible Interval

- Want an upper bound on λ

Summary: Bayesian Credible Interval

- Want an upper bound on λ
- A choice: find a λ_{max} such that

$$\int_0^{\lambda_{\text{max}}} \text{Posterior} = 90\%$$

Summary: Bayesian Credible Interval

- Want an upper bound on λ
- A choice: find a λ_{max} such that

$$\int_0^{\lambda_{\text{max}}} \text{Posterior} = 90\%$$

$$\Rightarrow \int_0^{\lambda_{\text{max}}} e^{-\lambda} = 0.9$$

$$\Rightarrow \lambda_{\text{max}} = \ln(10) \approx 2.3$$

Summary: Bayesian Credible Interval

- Want an upper bound on λ
- A choice: find a λ_{max} such that

$$\int_0^{\lambda_{\text{max}}} \text{Posterior} = 90\%$$

$$\Rightarrow \int_0^{\lambda_{\text{max}}} e^{-\lambda} = 0.9$$

$$\Rightarrow \lambda_{\text{max}} = \ln(10) \approx 2.3$$

Summary: Bayesian Credible Interval

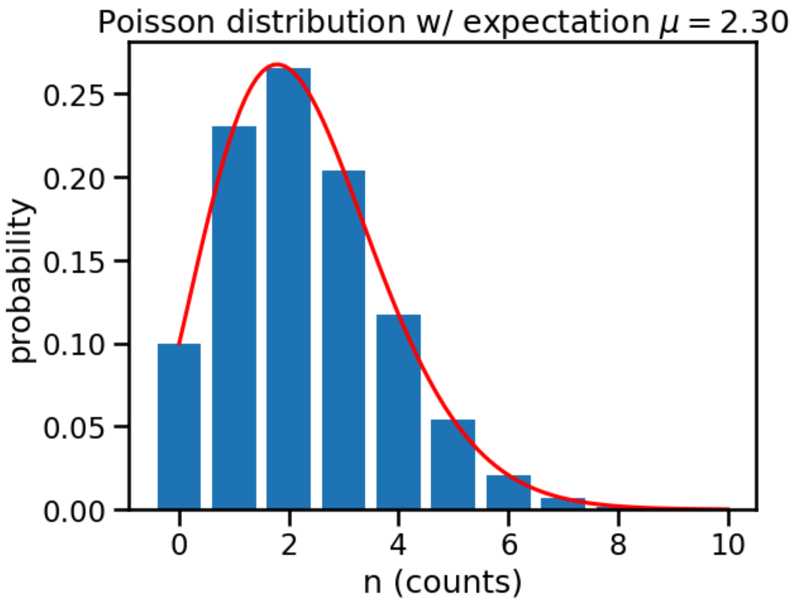
- Want an upper bound on λ
- A choice: find a λ_{max} such that

$$\int_0^{\lambda_{\text{max}}} \text{Posterior} = 90\%$$

$$\Rightarrow \int_0^{\lambda_{\text{max}}} e^{-\lambda} = 0.9$$

$$\Rightarrow \lambda_{\text{max}} = \ln(10) \approx 2.3$$

- so we "estemate with 90% confidence that $\lambda \leq 2.3$" base on a non-detection.



Summary: Bayesian Credible Interval

- Let μ denote the unknown parameter we wish to estimate.
- Let x_0 denote the outcome of a single measurement.
- Assume that we know how the measurement outcome depends on the parameter, $x = x(\mu)$.
  - *e.g.* if the neutrino flux is very small, then oftentimes a measurement reports a non-detection.
  - In other words, we know the *likelihood*, $P(x_0|\mu)$.
- From the Bayesian perspective, we can flip things around and say that the parameter is a function of the measurement, $P(\mu|x_0)$, provided that we state our prior beliefs about the parameter, $P(\mu)$.