90% Confidence Level Upper Bound

Brief discussion of Feldman & Cousins

10-31-2024

Outline

Parameter Estimataion

Poisson Distribution

- Expect an average of μ counts per time, with $\mu \in \mathbb{R} \geq 0$

Poisson Distribution

- Expect an average of μ counts per time, with $\mu \in \mathbb{R} \geq 0$
- Probability of getting $n \in \mathbb{N}$ counts per time?

Poisson Distribution

- Expect an average of μ counts per time, with $\mu \in \mathbb{R} \geq 0$
- Probability of getting $n \in \mathbb{N}$ counts per time?
- Poisson Distribution:

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

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

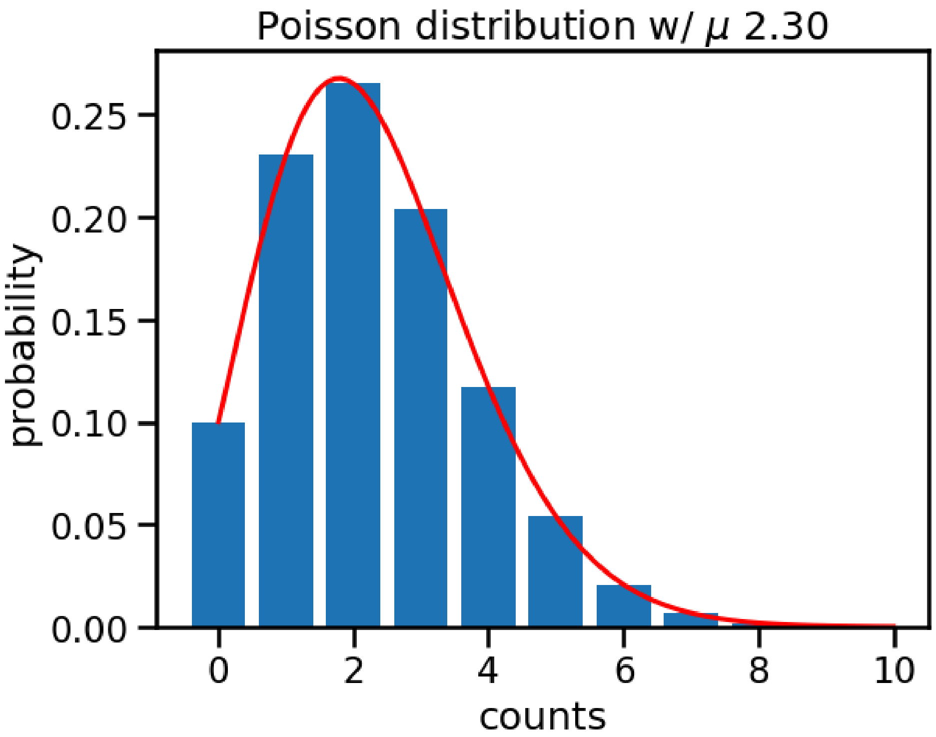
Poisson Distribution

- Expect an average of μ counts per time, with $\mu \in \mathbb{R} \geq 0$
- Probability of getting $n \in \mathbb{N}$ counts 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; suppose the first measurement yields x_0.

CI Definition

- Goal: estimate parameter μ whose true value is μ_t
- Make a measurement x; suppose the 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; suppose the 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

CI Definition

- Goal: estimate parameter μ whose true value is μ_t
- Make a measurement x; suppose the 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

CI Definition

- Goal: estimate parameter μ whose true value is μ_t
- Make a measurement x; suppose the 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 $[\mu_l(x_1), \mu_u(x_1)]$

CI Definition

- Goal: estimate parameter μ whose true value is μ_t
- Make a measurement x; suppose the 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 $[\mu_l(x_1), \mu_u(x_1)]$
- 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

- The set C has the property that

$$P([\mu_l, \mu_u] \ni \mu) = \alpha \qquad \forall \text{ allowed } \mu$$

CI Definition

- The set C has the property that

$$P([\mu_l, \mu_u] \ni \mu) = \alpha \qquad \forall \text{ allowed } \mu$$

- In words:
  - pick a *fixed* value of μ; say we pick the truth μ_t.

CI Definition

- The set C has the property that

$$P([\mu_l, \mu_u] \ni \mu) = \alpha \qquad \forall \text{ allowed } \mu$$

- In words:
  - pick a *fixed* value of μ; say we pick the truth μ_t.
  - of the members in C, a fraction α would cover μ_t.

CI Definition

- The set C has the property that

$$P([\mu_l, \mu_u] \ni \mu) = \alpha \qquad \forall \text{ allowed } \mu$$

- In words:
  - pick a *fixed* value of μ; say we pick the truth μ_t.
  - of the members in C, a fraction α would cover μ_t.
  - others won't.

CI Definition

- The set C has the property that

$$P([\mu_l, \mu_u] \ni \mu) = \alpha \qquad \forall \text{ allowed } \mu$$

- In words:
  - pick a *fixed* value of μ; say we pick the truth μ_t.
  - of the members in C, a fraction α would cover μ_t.
  - others won't.
- The members of C are called *confidence intervals*.

CI Construction: Confidence Belt

- Consider probability
  $\Pr(x|\mu)$

CI Construction: Confidence Belt

- Consider probability $\Pr(x|\mu)$
- Take for example $\mu = 4$, consider $\Pr(x|\mu = 4)$

CI Construction: Confidence Belt

- Consider probability $\Pr(x|\mu)$
- Take for example $\mu = 4$, consider $\Pr(x|\mu = 4)$
- 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

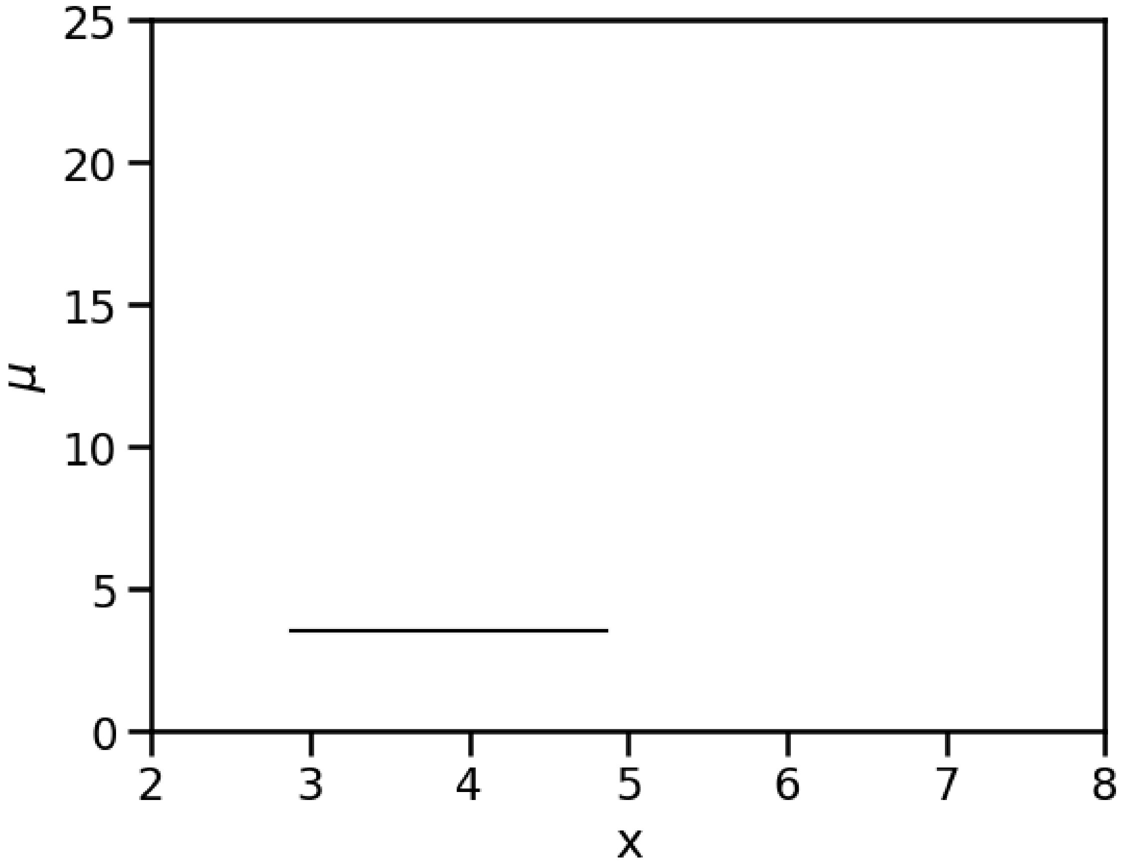
- Consider probability $\Pr(x|\mu)$
- Take for example $\mu = 4$, consider $\Pr(x|\mu = 4)$
- Select a region $[x_l, x_u]$ such that the probability of measuring $x \in [x_l, x_u]$ is, say, 80%.
- that is,

$$\Pr(x \in [x_l, x_u]|\mu = 4) = 80\%$$

CI Construction: Confidence Belt

- Consider probability $\Pr(x|\mu)$
- Take for example $\mu = 4$, consider $\Pr(x|\mu = 4)$
- Select a region $[x_l, x_u]$ such that the probability of measuring $x \in [x_l, x_u]$ is, say, 80%.
- that is,

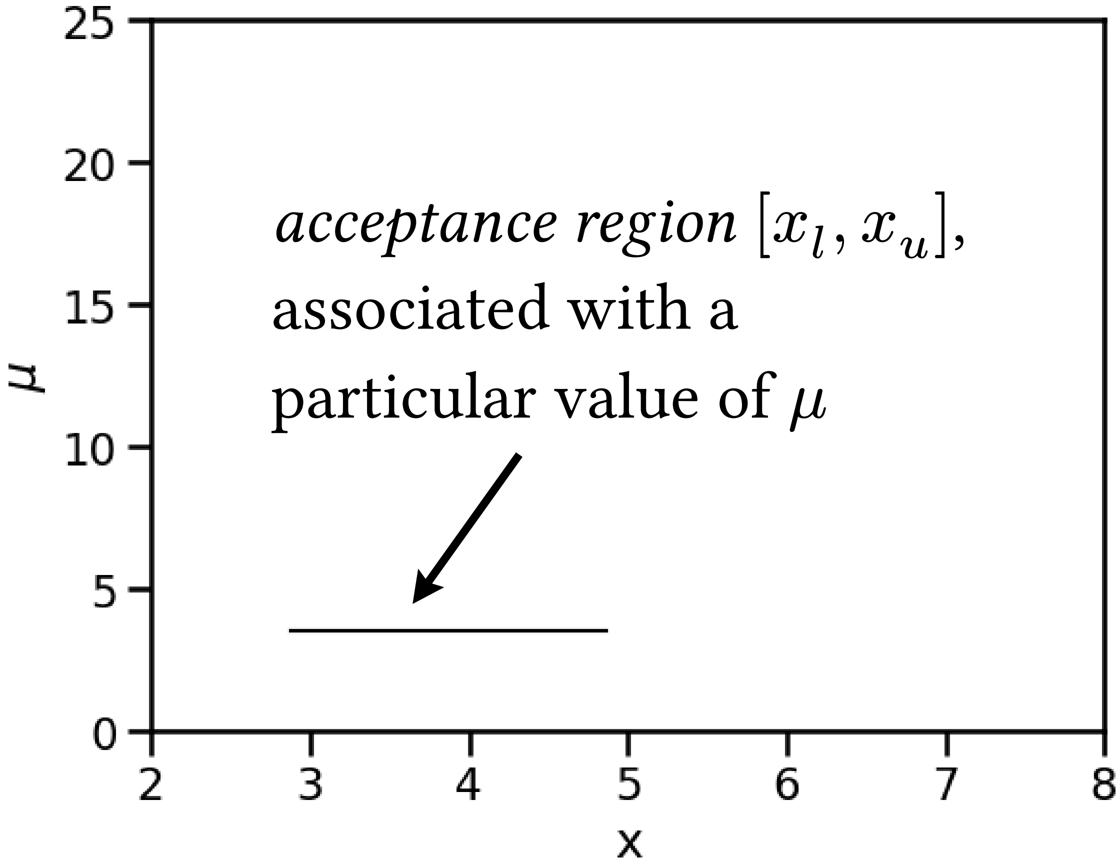
$$\Pr(x \in [x_l, x_u]|\mu = 4) = 80\%$$



CI Construction: Confidence Belt

- Consider probability $\Pr(x|\mu)$
- Take for example $\mu = 4$, consider $\Pr(x|\mu = 4)$
- Select a region $[x_l, x_u]$ such that the probability of measuring $x \in [x_l, x_u]$ is, say, 80%.
- that is,

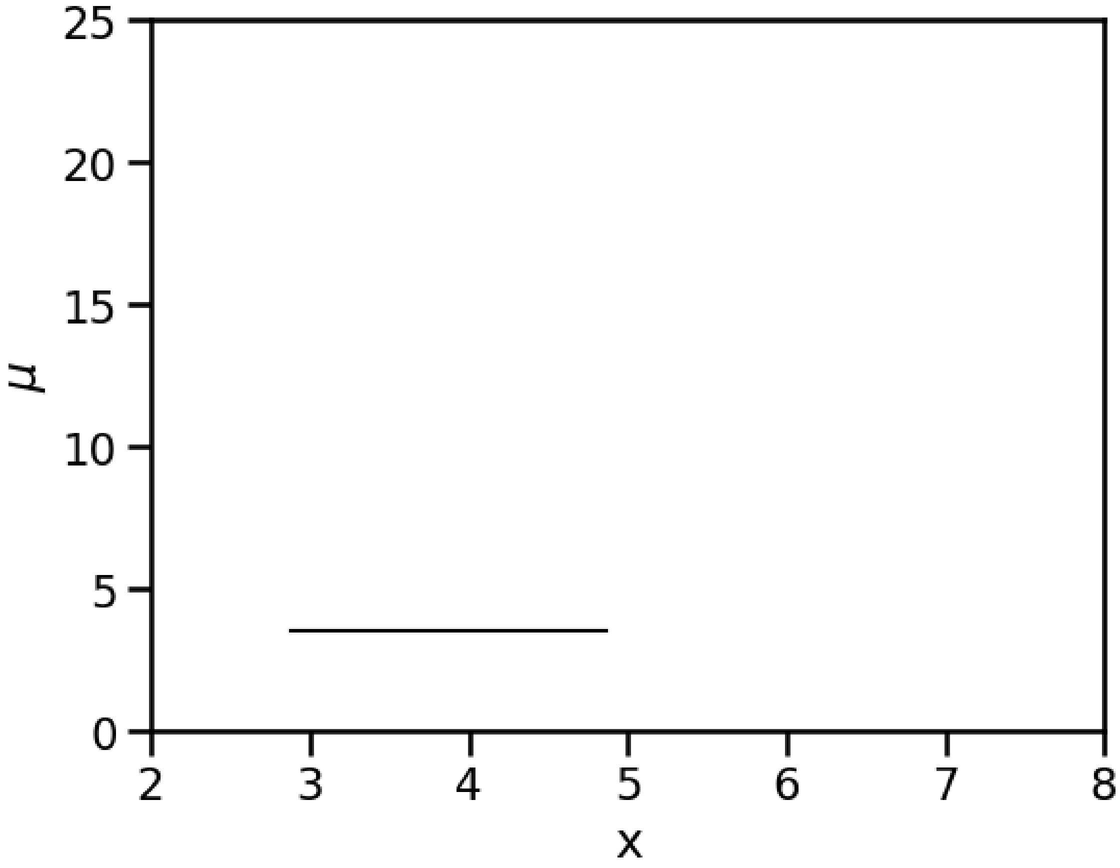
$$\Pr(x \in [x_l, x_u] | \mu = 4) = 80\%$$



CI Construction: Confidence Belt

- Take another value μ, say $\mu = 5$
- Get another acceptance region $[x_l, x_u]$ associated with this value

$$\Pr(x \in [x_l, x_u] | \mu = 5) = 80\%$$

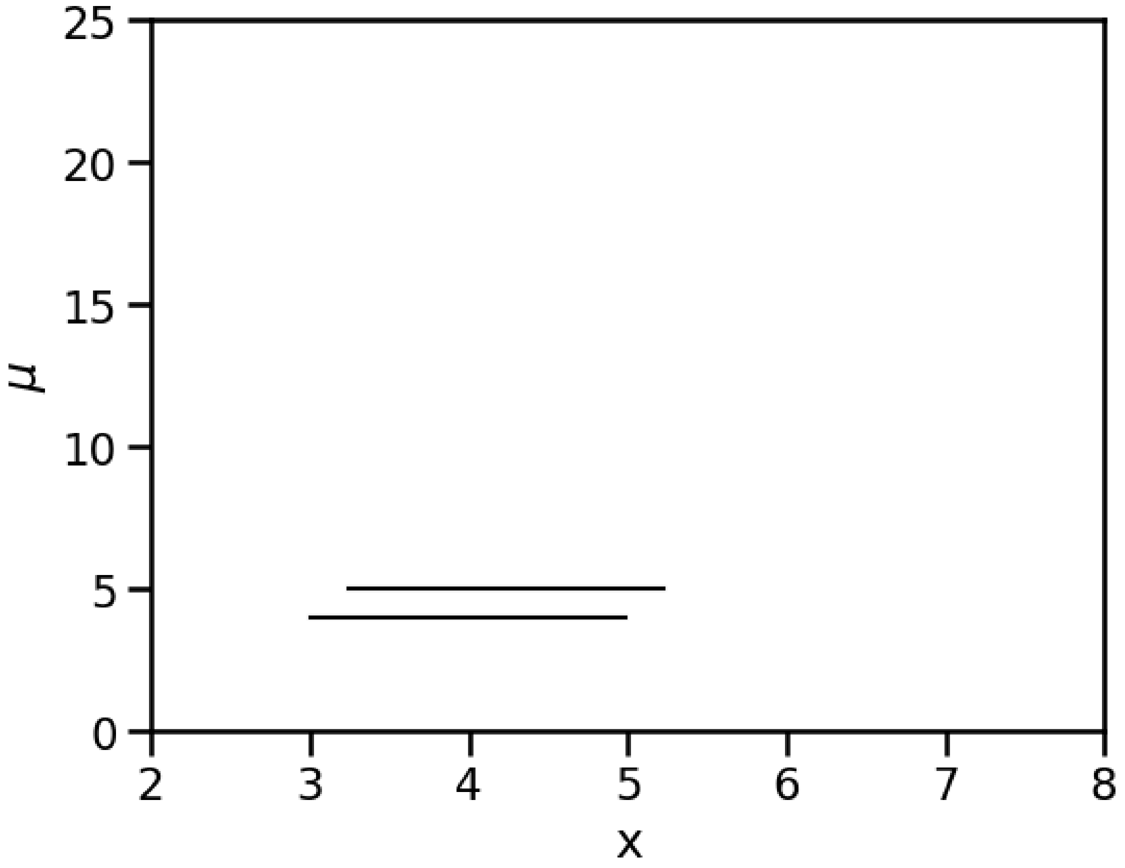


CI Construction: Confidence Belt

- Take another value μ, say $\mu = 5$
- Get another acceptance region $[x_l, x_u]$ associated with this value

$$\Pr(x \in [x_l, x_u] | \mu = 5) = 80\%$$

- Rinse and repeat

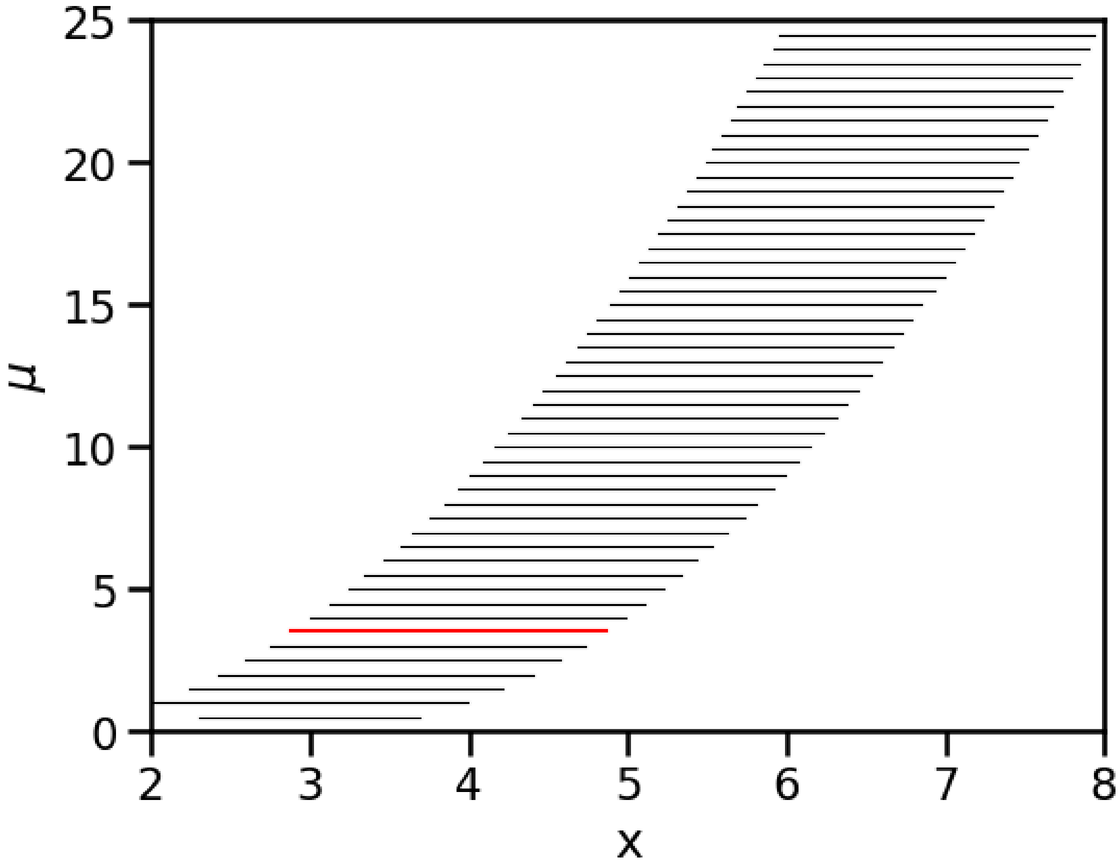


CI Construction: Confidence Belt

- Take another value μ, say $\mu = 5$
- Get another acceptance region $[x_l, x_u]$ associated with this value

$$\Pr(x \in [x_l, x_u] | \mu = 5) = 80\%$$

- Rinse and repeat

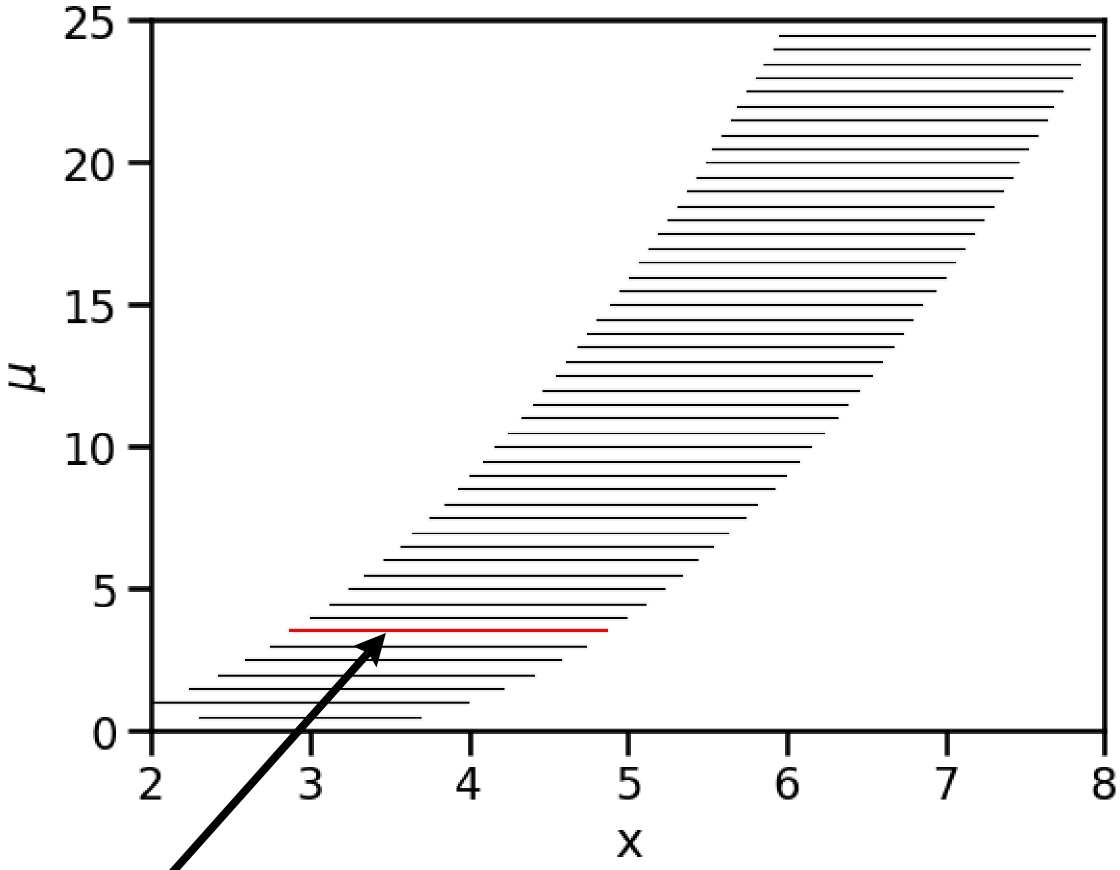


CI Construction: Confidence Belt

- Take another value μ, say $\mu = 5$
- Get another acceptance region $[x_l, x_u]$ associated with this value

$$\Pr(x \in [x_l, x_u] | \mu = 5) = 80\%$$

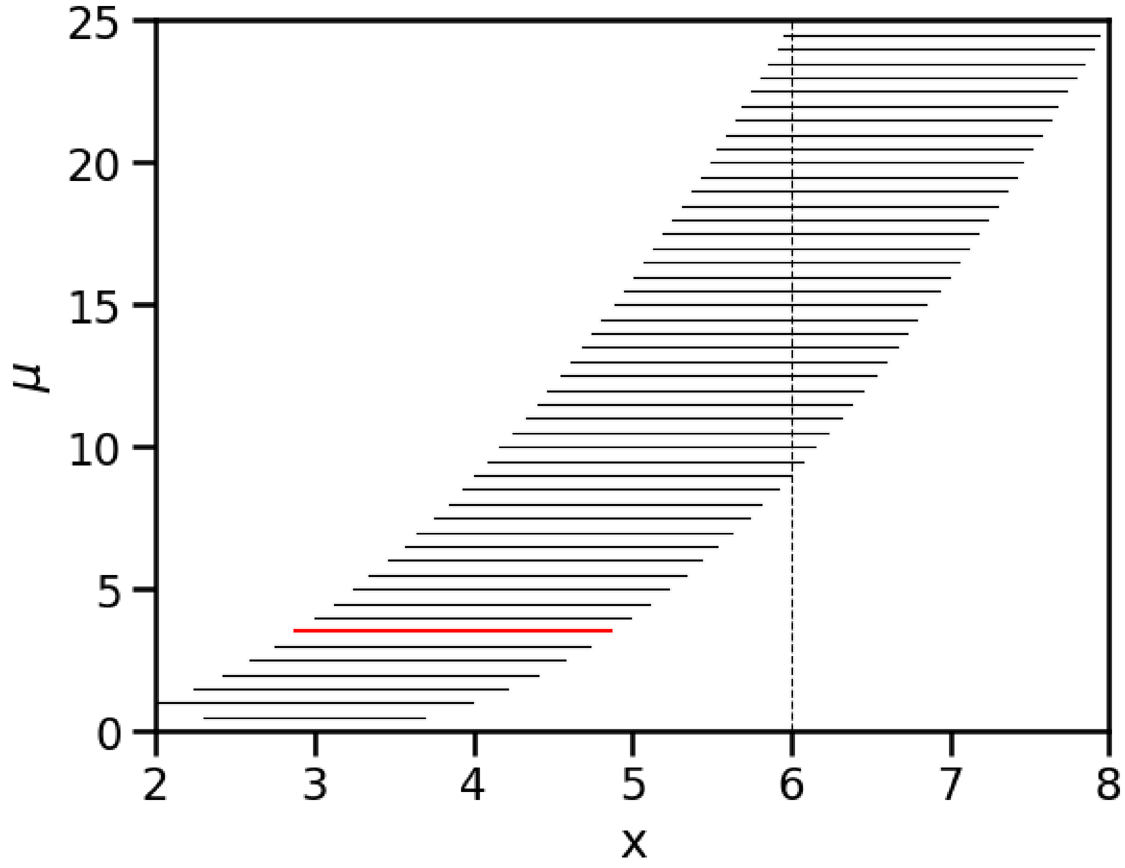
- Rinse and repeat



suppose this is μ_t

CI Construction: Confidence Belt

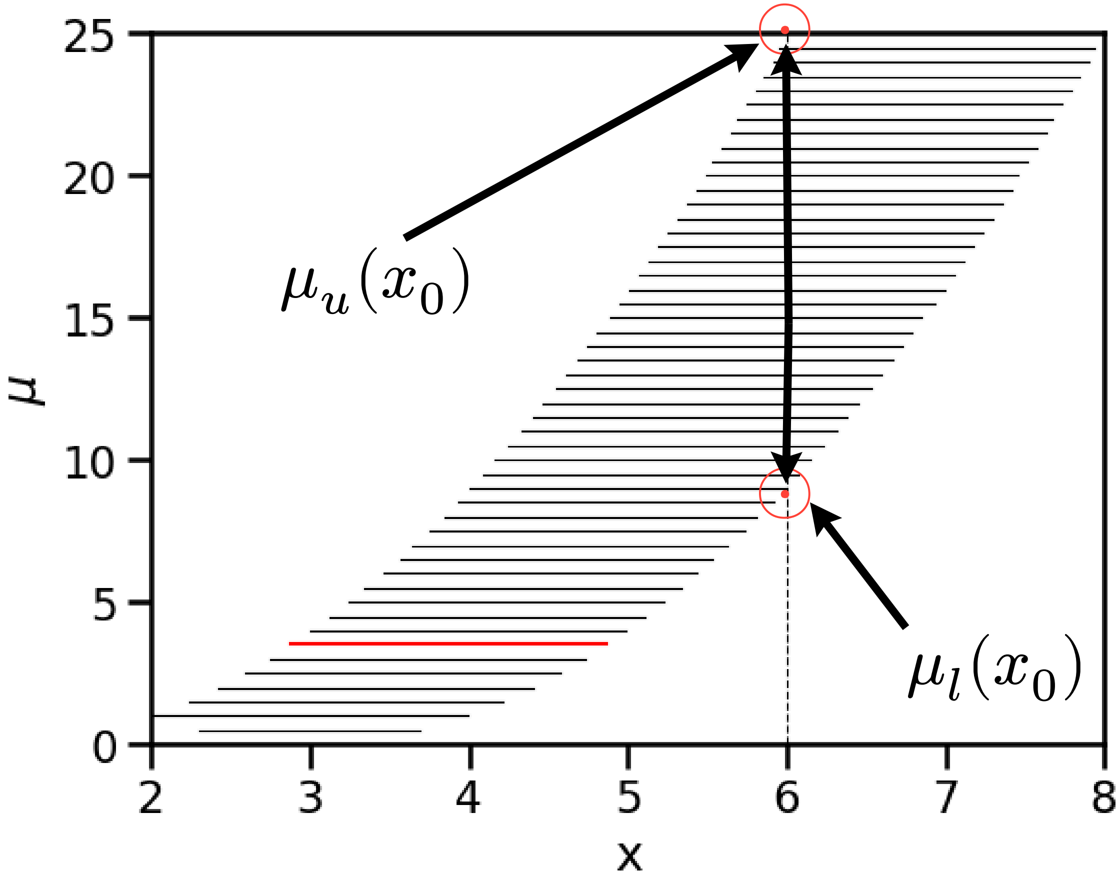
- Make a measurement, get result x_0
- The probability of x_0 falling in the region is 80%, by construction



CI Construction: Confidence Belt

- Make a measurement, get result x_0
- The probability of x_0 falling in the region is 80%, by construction
- The **confidence interval** $[\mu_l, \mu_u]$ from this experiment is

the vertical intercept.

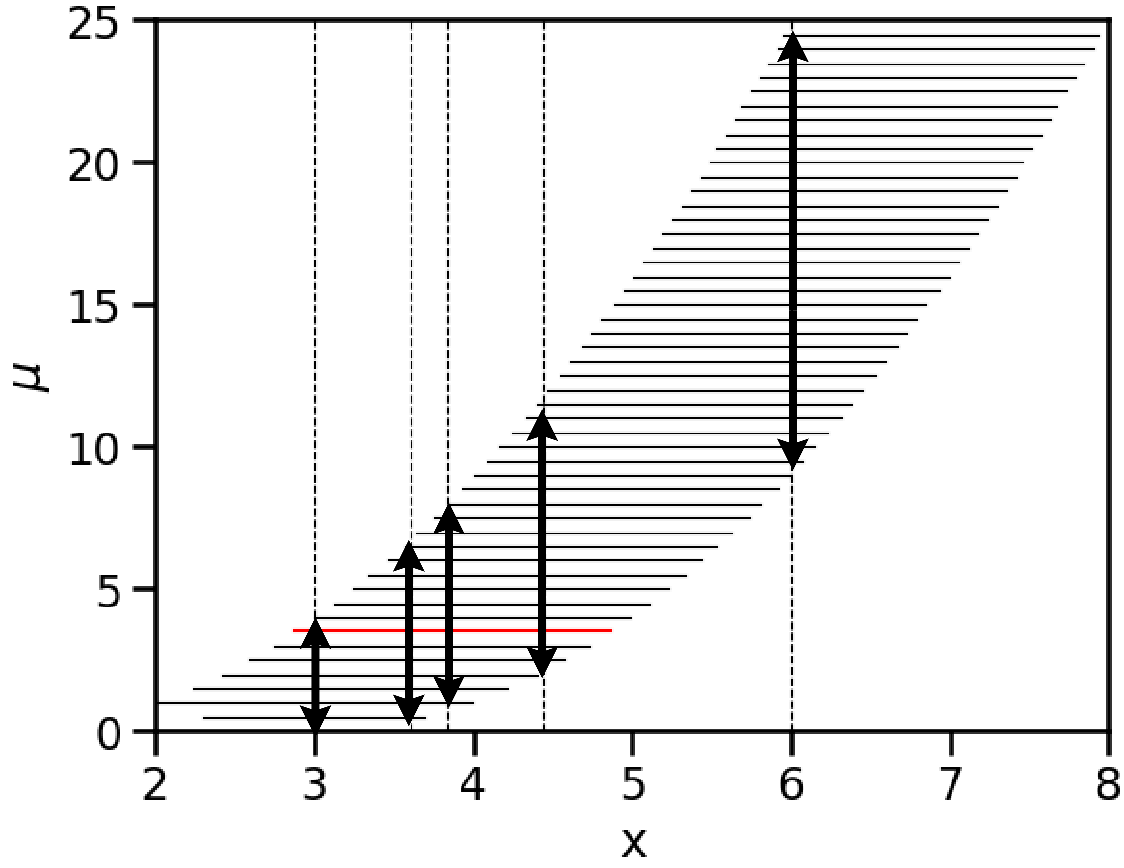


CI Construction: Confidence Belt

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

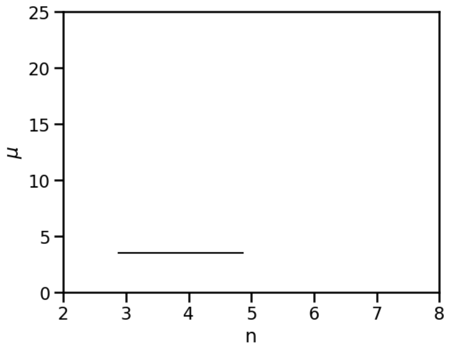
$$C = \{\mathrm{CI}_1, \mathrm{CI}_2, \mathrm{CI}_3, \mathrm{CI}_4, \mathrm{CI}_5\}$$

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



Acceptance Region

Maximum Likelihood



- Recall acceptance region:

$$\Pr(n \in \boxed{[n_1, n_2]} \mid \mu_{\text{fixed}}) = 80\%$$

- Complete freedom in choosing how to construct the acceptance regions.

- Consider 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" $\boxed{\mu_{\text{best}} \equiv \max(0, n - b)}$

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	

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

$$\underbrace{\Pr(\lambda \mid n)}_{\textit{Posterior}} = \frac{\Pr(n \mid \lambda) \cdot \Pr(\lambda)}{\Pr(n)}$$

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

$$\underbrace{\Pr(\lambda \mid n)}_{\text{Posterior}} = \frac{\Pr(n \mid \lambda) \cdot \Pr(\lambda)}{\Pr(n)}$$

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

$$\underbrace{\Pr(\lambda \mid n)}_{\textit{Posterior}} = \frac{\overbrace{\Pr(n \mid \lambda)}^{\textit{Likelihood}} \cdot \Pr(\lambda)}{\underbrace{\Pr(n)}_{\textit{evidence}}}$$

- 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

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

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

Bayes Theorem

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

- 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" $\Pr(\lambda) = 1 \implies \Pr(\lambda \mid n) = \Pr(n \mid \lambda)$

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

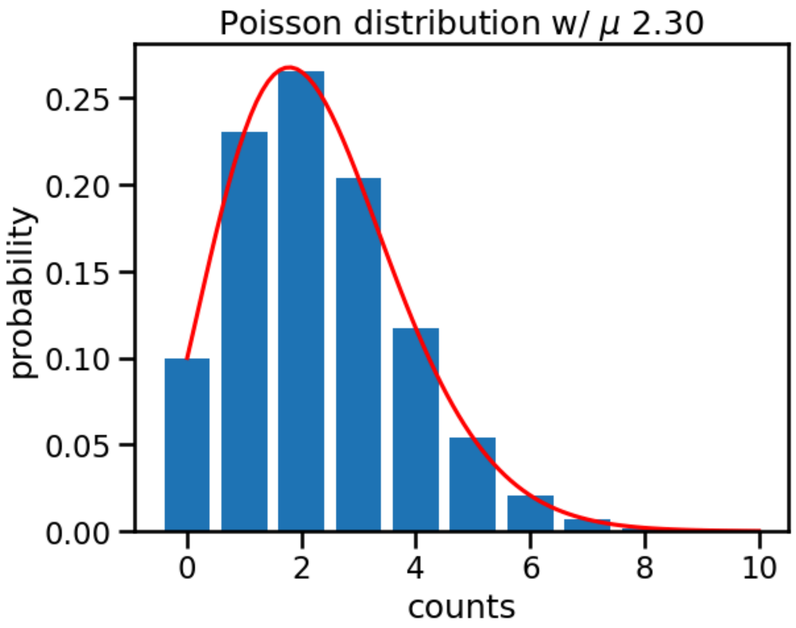
- Want an upper bound on λ
- A choice: find a λ_{max} such that

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

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

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

- 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, $P(\mu)$.