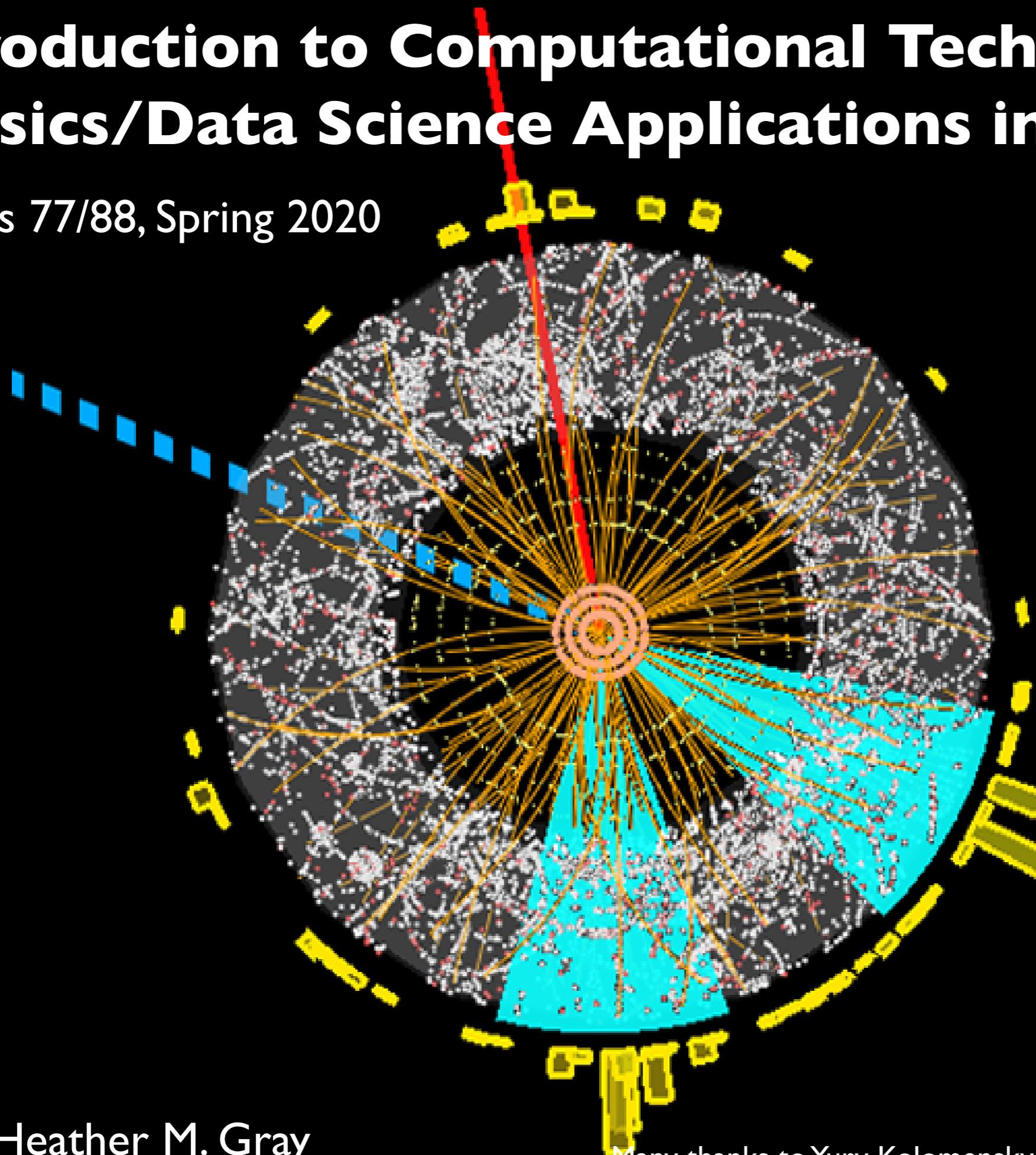


Introduction to Computational Techniques in Physics/Data Science Applications in Physics

Physics 77/88, Spring 2020



**Statistics and
Probability,
Interpreting
Measurements**

There are three kinds of lies: lies, damned lies and statistics
- Mark Twain

Announcements

- Grades for HW1 have been released — solutions are available on bcourses
 - Remember the honor code — important to keep the solutions confidential
- **Reminder: Regrading policy**
 - If you think that a problem needs regrading, please write a paragraph explaining the issue and why it should be regraded, please send the paragraph as an attachment in an **email** to me and Nick. Please note that if your problem is regraded, I will regrade the full homework and your grade may go up or down. Please note that this will be **the only method** by which to request regrading
- Please remember to fill in the survey for each lecture
 - Regular participation counts towards class participation
- Office hours (check my calendar in general)
 - This week: Monday 12 - 1 pm
 - Next week: Monday 10 - 11 am

The Statistics Boot Camp (Lectures 5 and 6)

- Introduction
- Definitions: results of experiments
 - Random variables, probability, PDFs
- Interpreting results
 - Point estimators
 - Max likelihood, least squares fits
- Hypothesis testing, confidence limits
- Simulation (Monte Carlo techniques) — in two weeks

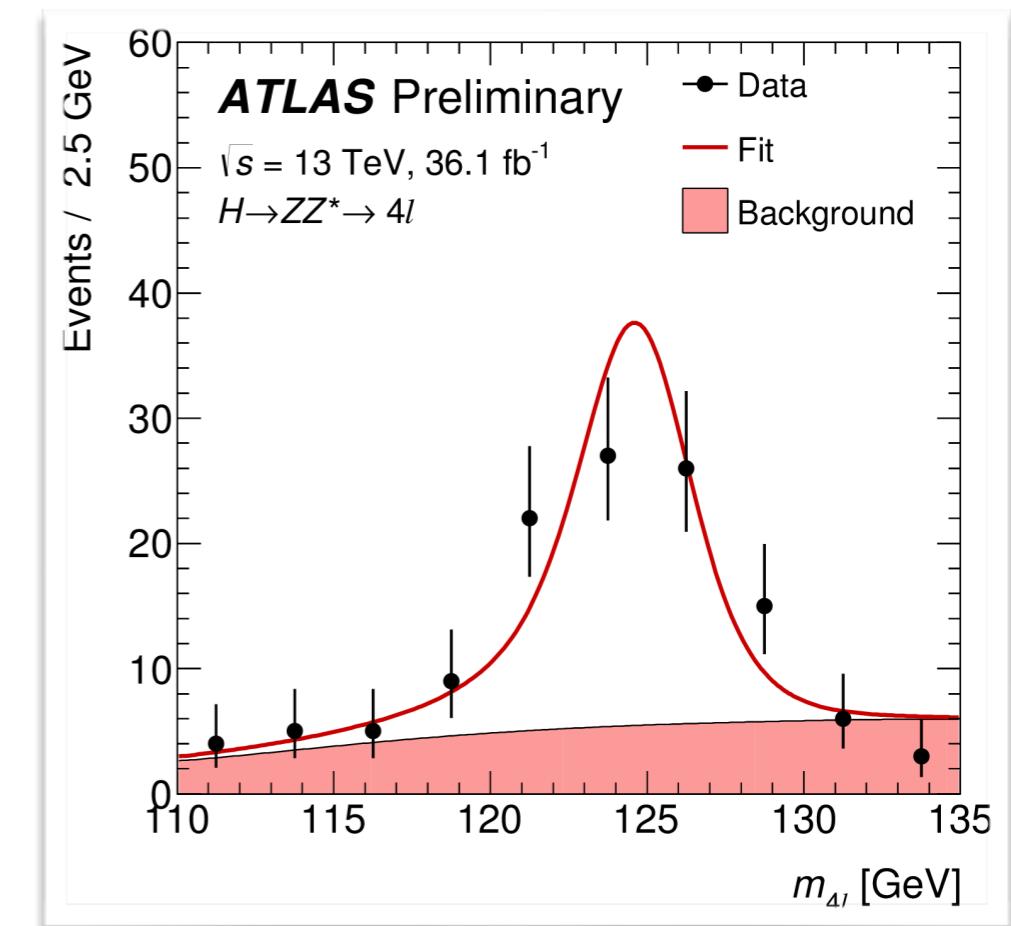
Statistics

- Statistics is a vital tool
 - Physics is an experimental science
 - Requires both and understanding
- From to
 - Make a set of
 - Summarise the
 - Most conclusions are drawn with some degree of
- Example
 -
 -
 -
- Many measurements are a priori uncertain and have to be interpreted in probabilistic terms
- **Classical statistics:** estimate and test whether a given is given a with the data

This may seem rather dry, but if you turn out to be an experimentalist it's really important and useful!

Describing the Data

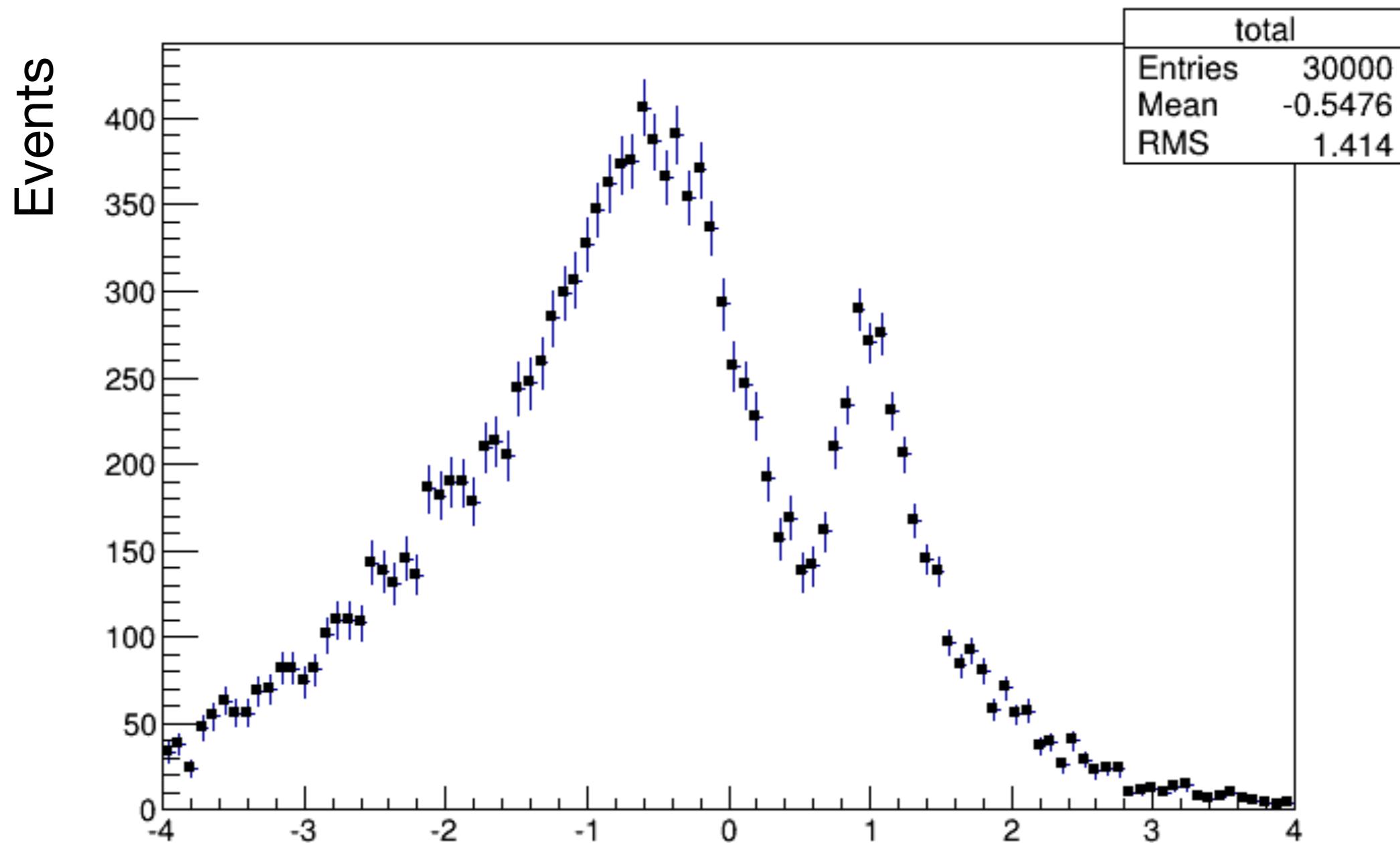
- **Data:**
 - Physics:
 - Other fields deal with
 - **Numbers** are easier to handle mathematically; statistics will deal with quantitative measurements
 - data, e.g. integers (counts)
 - data, e.g. energies, momenta
 - Measure with some , set by the measuring apparatus or other external conditions



Why are there error bars on the data?

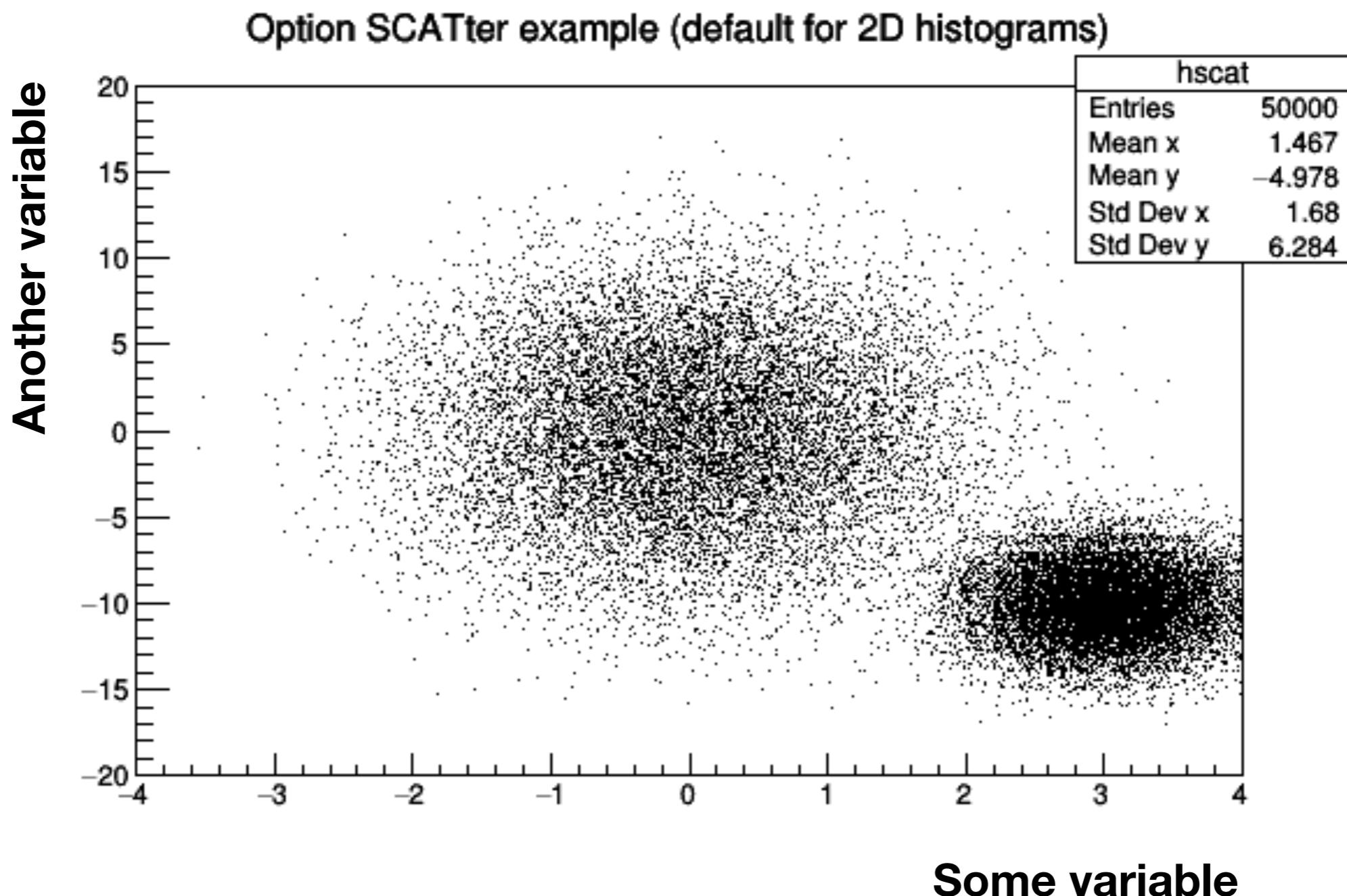
Histogram

This is the total distribution



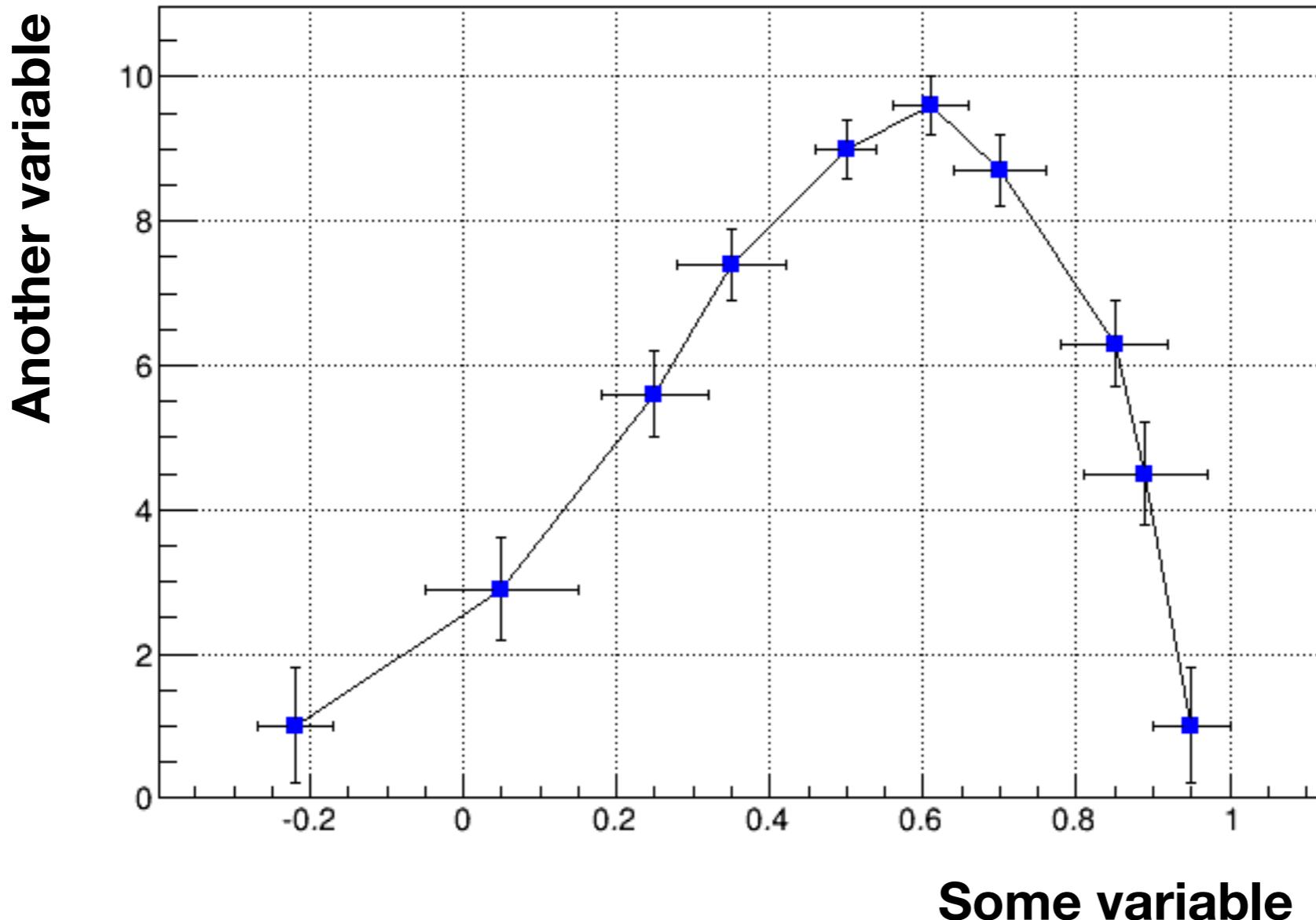
Some variable

Scatterplots



Graph

TGraphErrors Example



Uncertainty and Error

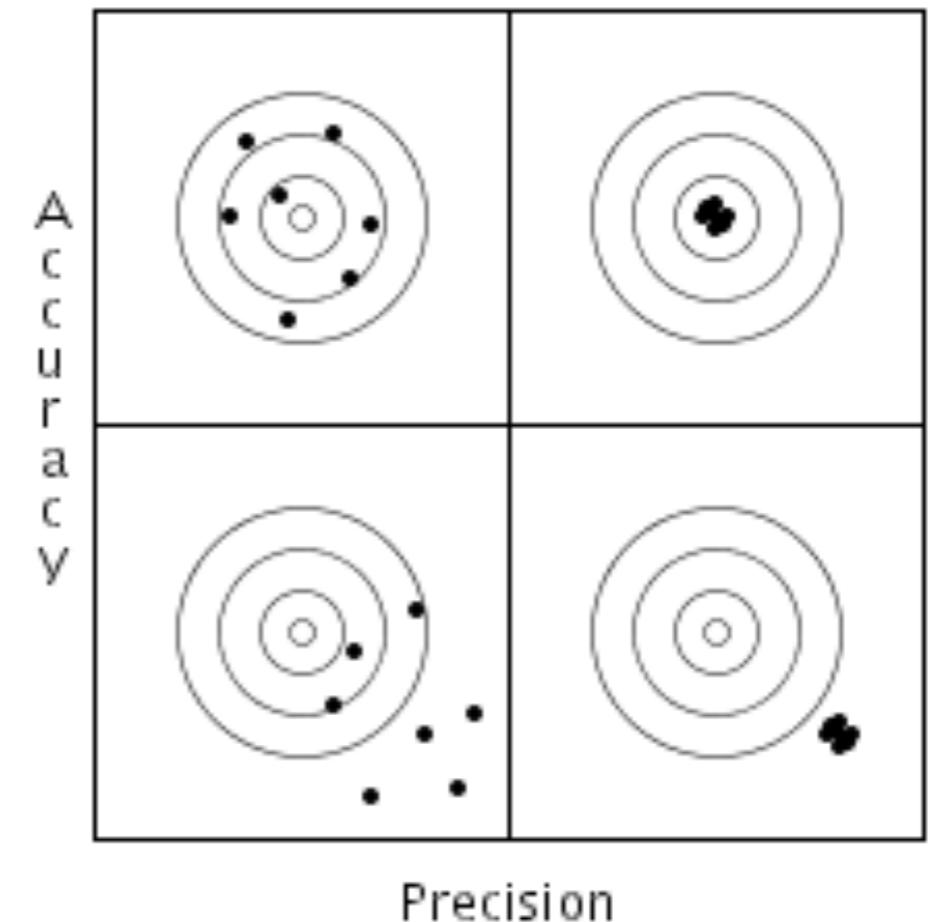
- In physics, the words and are used interchangeably to describe how far a measurement — typically
 - Use symbol σ for the standard deviation
 - Formal definition is probabilistic: 68% of measurements fall within 1σ of the mean
- Often interpreted as a range of true values
- We'll come back to the difference between measured and true values

Uncertainty and Error

- How do we what is ?
 - Underlying assumption: our experiment is of a of
 - Derive the from the of the
 - Implicit assumption: our experiment is
 - i.e. all experiments would return results

Precision vs Accuracy

- Precision: of the data around the value
 - Typically associated with uncertainty
- Accuracy: of the value from the value
 - Typically associated with uncertainty
- Bad data:
 - Data with distribution (e.g.)



http://anomaly.org/wade/blog/2006/01/accuracy_and_precision.html

Golden Rules

- When reporting **of a measurement**, report its **uncertainty**
- Round off values to **3 significant digits**
 - Rule of thumb: **if the last digit is 5 or greater, round up**, otherwise **round down**
 - $x = 12.345 \pm 0.024$
 - $y = 12.3456 \pm 0.6$
- Uncertainty can come from the **accuracy** and/or **precision** of the measurement
 - Rule of thumb: **use the smaller uncertainty**
 - Statistically correct:

Probability: Definitions

- For numerical data, is often most convenient (and quantitative)
- Let's define probability now
 - Formally, it is a that is defined by
 - 1.
 - 2.
 - 3.

Two Interpretations

- “Frequentist” interpretation:
 - Probability is a measure of how often a given parameter value is reported when experiments are repeated many times
 - Measurable parameters are represented by random variables assigned to the outcomes of experiments
 - CL measures a confidence interval, given a sample size, such that an estimator would fall in a certain range a specified percentage of the time
 - No probability is assigned to individual parameter values
- “Bayesian” interpretation:
 - More general: define probability as a measure of the degree of belief that a given statement is true
 - E.g. that the posterior probability of a parameter value being in a certain interval is a given percentage
 - This is somewhat subjective, but follows how people actually reason

Frequentist Probability

- Definitions
 - Let Ω be the sample space of all possible outcomes of an experiment.
 - Any subset $A \subseteq \Omega$ with only finitely many elements is called an event.
 - Define $P(A) = \frac{|A|}{|\Omega|}$
 - Assume that the outcomes in Ω are (in principle) equally likely.
 - Confidence in a hypothesis grows with the number of observations.
 - Frequentist statistics is argued for) in situations where the validity of statistical methods can be tested by repeated sampling.
 - (e.g. particle physics)



John von Neumann



Jerzy Neyman

Bayes Theorem

- probability of given
-
- Interpreted within as
 - $P(\text{theory} | \text{data}) \propto$
- Allows one to a as a measure of that a given is correct (e.g. that some is in some)
- Requires assigning some to prior knowledge
 - That's where comes in



Thomas Bayes

Probability: Random Variables and PDFs

- For a continuous variable, x , we define the

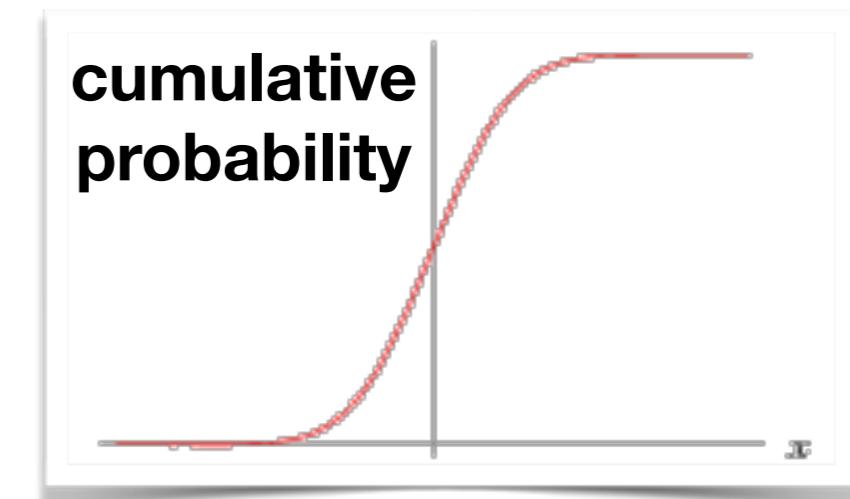
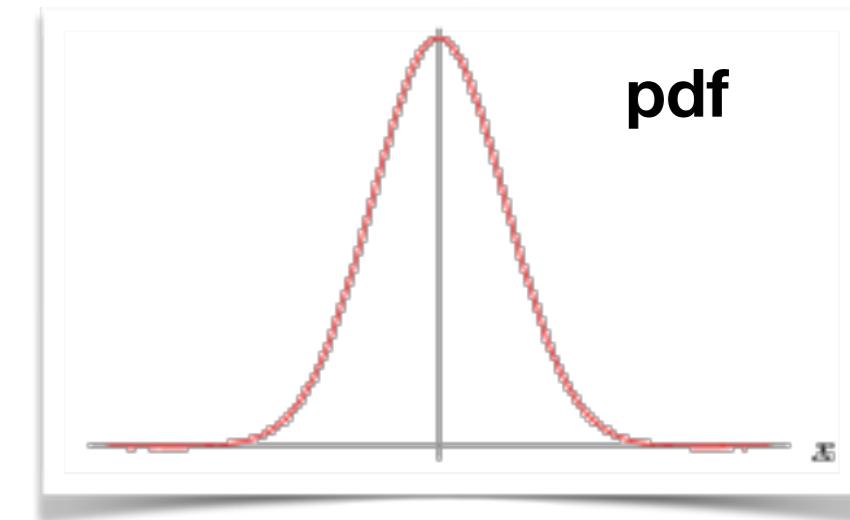
- $f(x, \theta) =$

- θ

- Integrate to obtain the

- $F(a) = \int$

- Probability that $x < a$
- Discrete variables:
- Expectation value



- $E[u(x)] =$

Expectation Values

- Expectation value of a

- $E[u(x)] =$

- Moments of a

- $\alpha_n \equiv$

- $m_n \equiv$

Mean and Variance

- Mean

- $\mu = \int$

- Variance

- $\sigma^2 =$

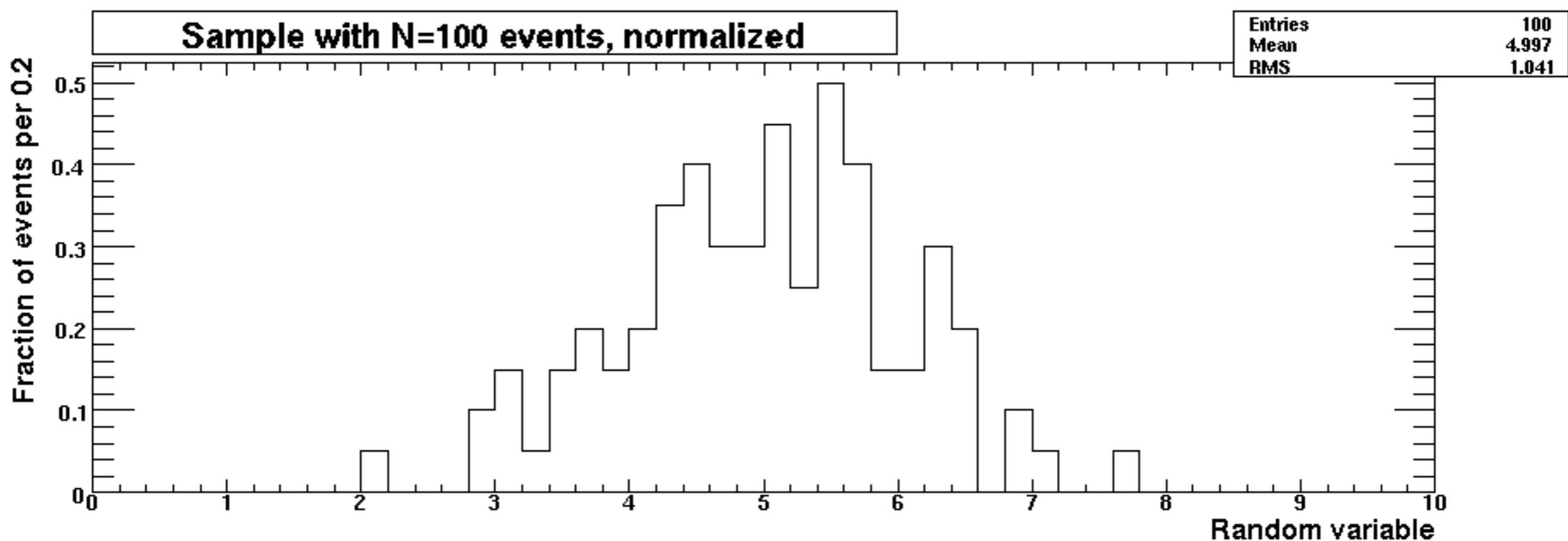
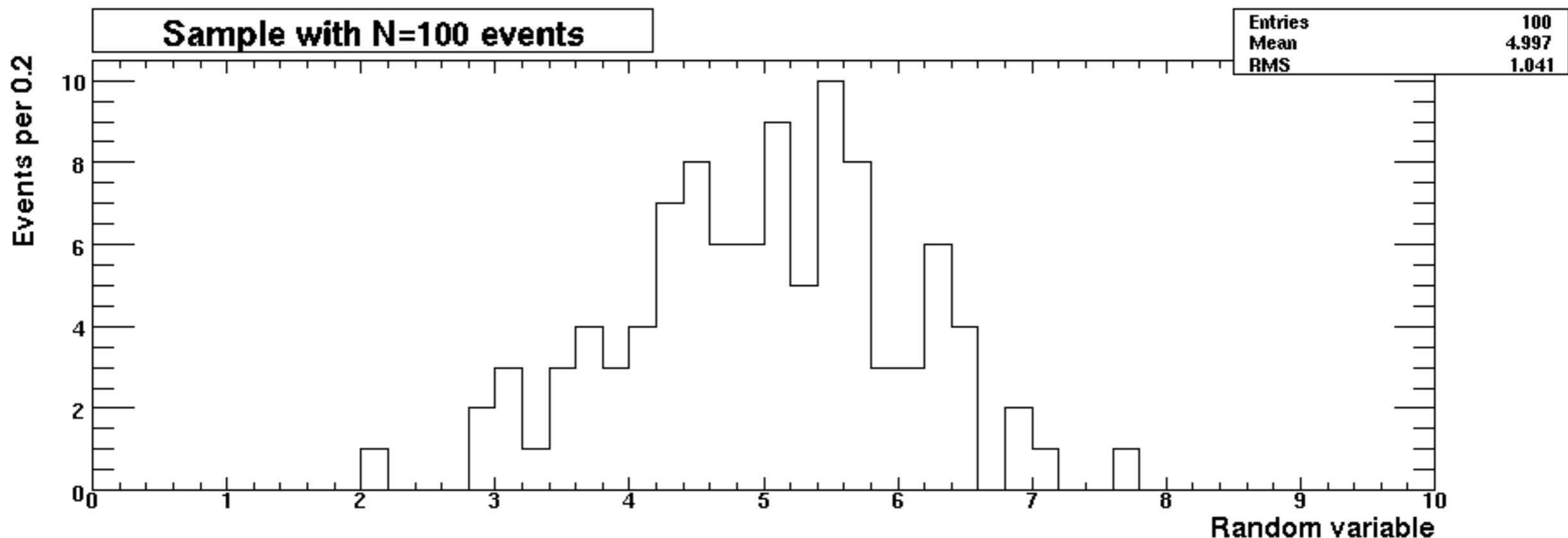
- σ is the standard deviation

If you only remember one thing today, remember this

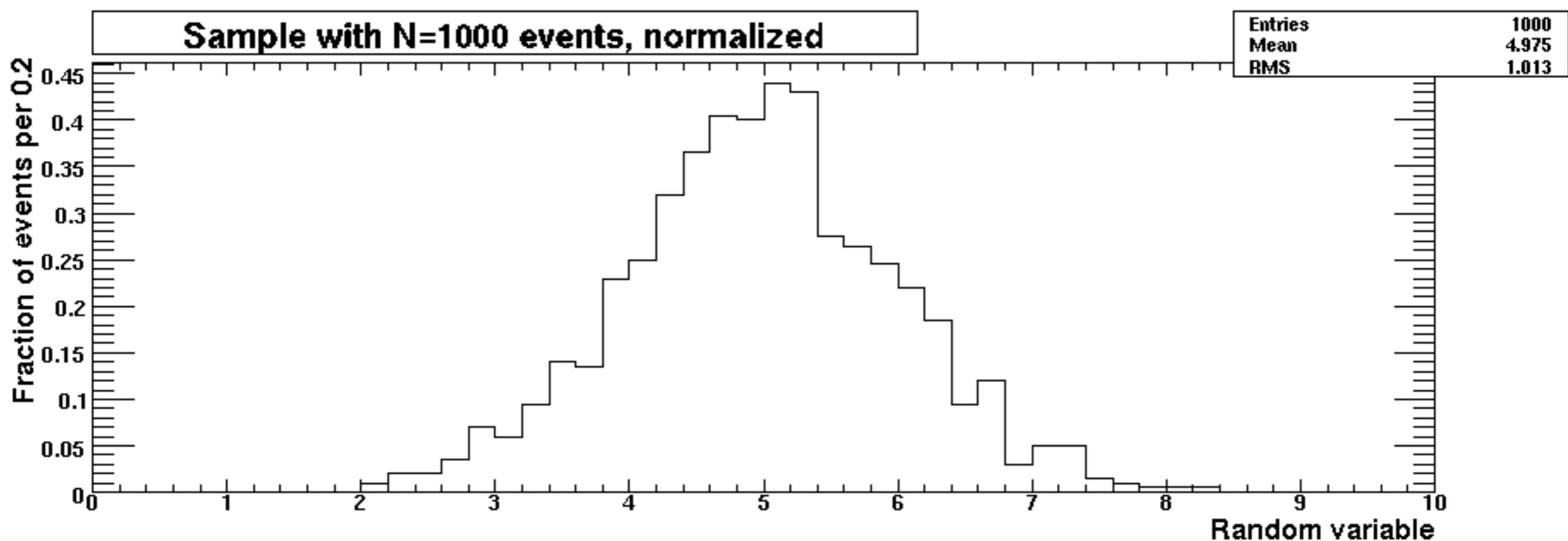
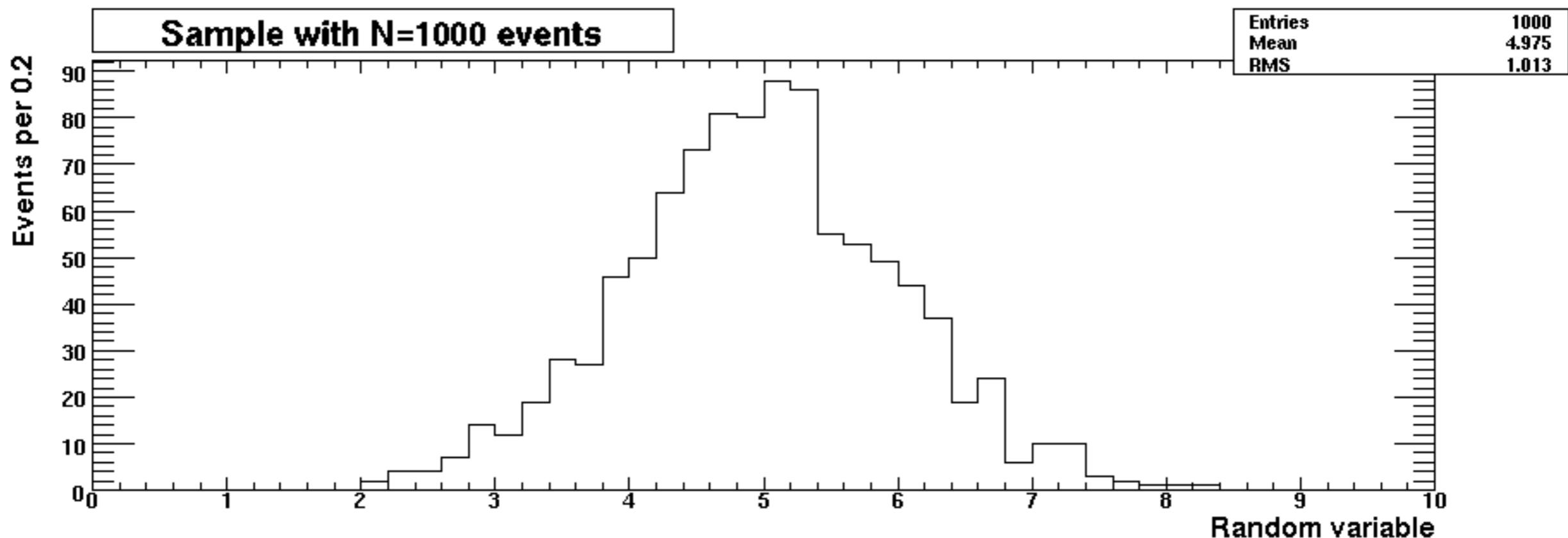
If we know the PDF, we know how to determine μ and σ

Example in jupiter notebook

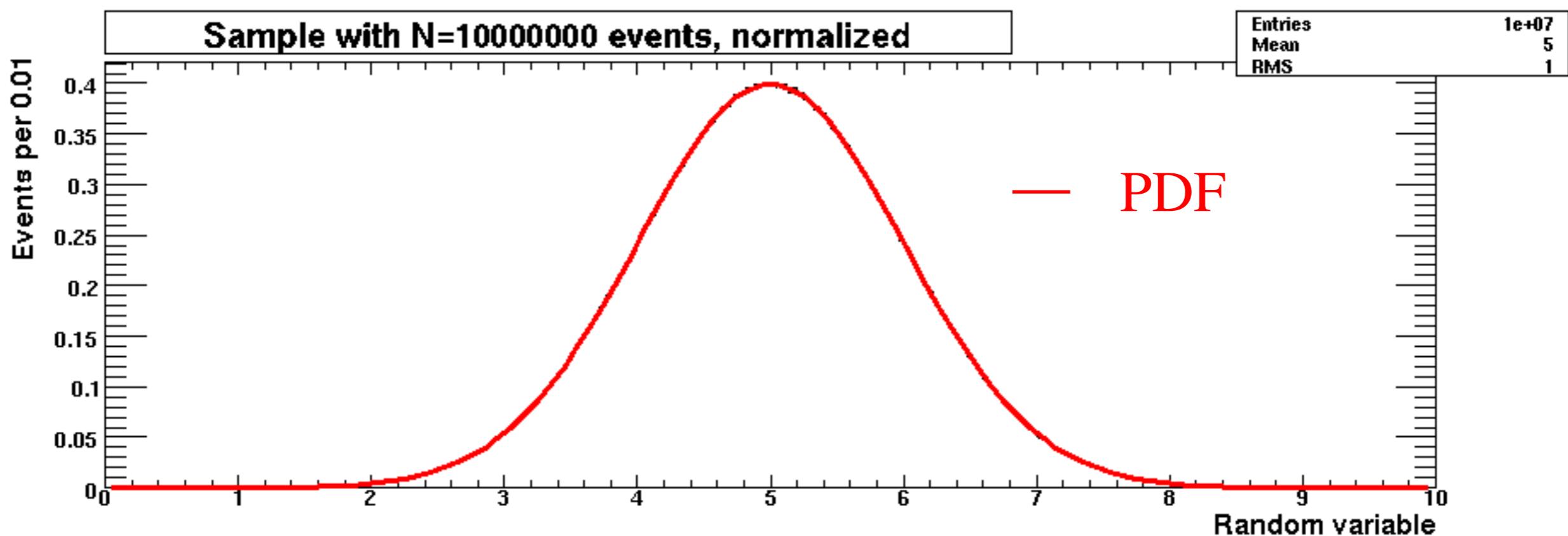
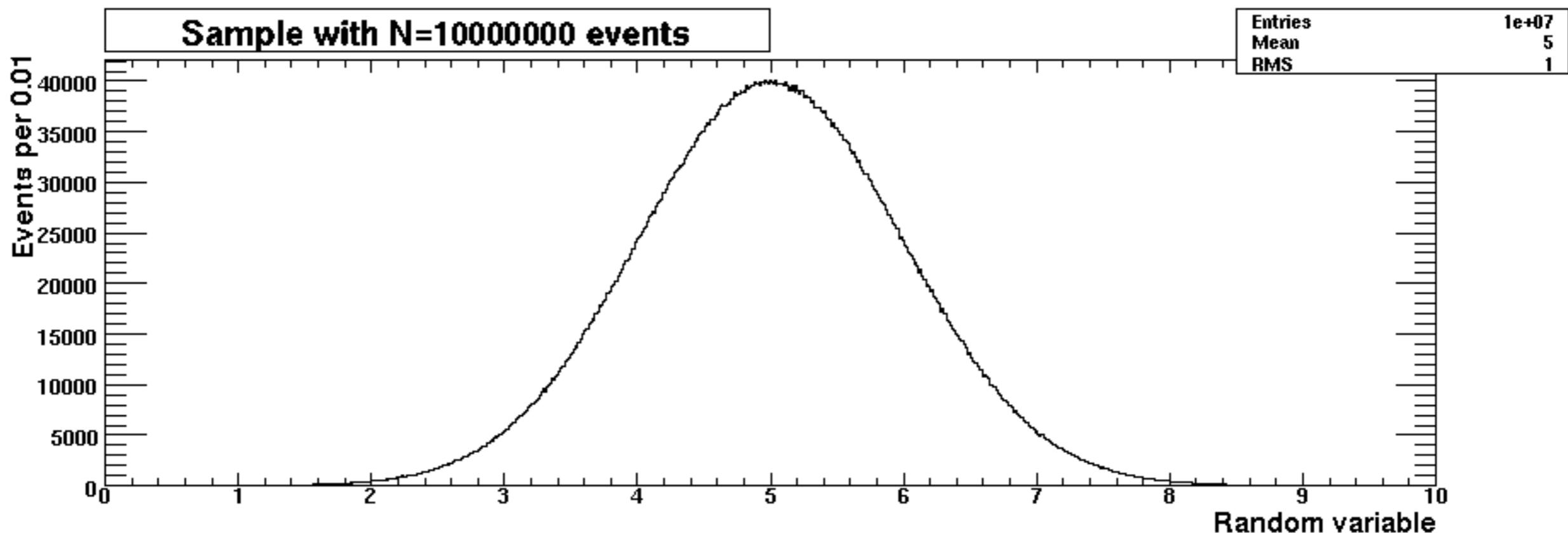
Sample from a Continuous PDF



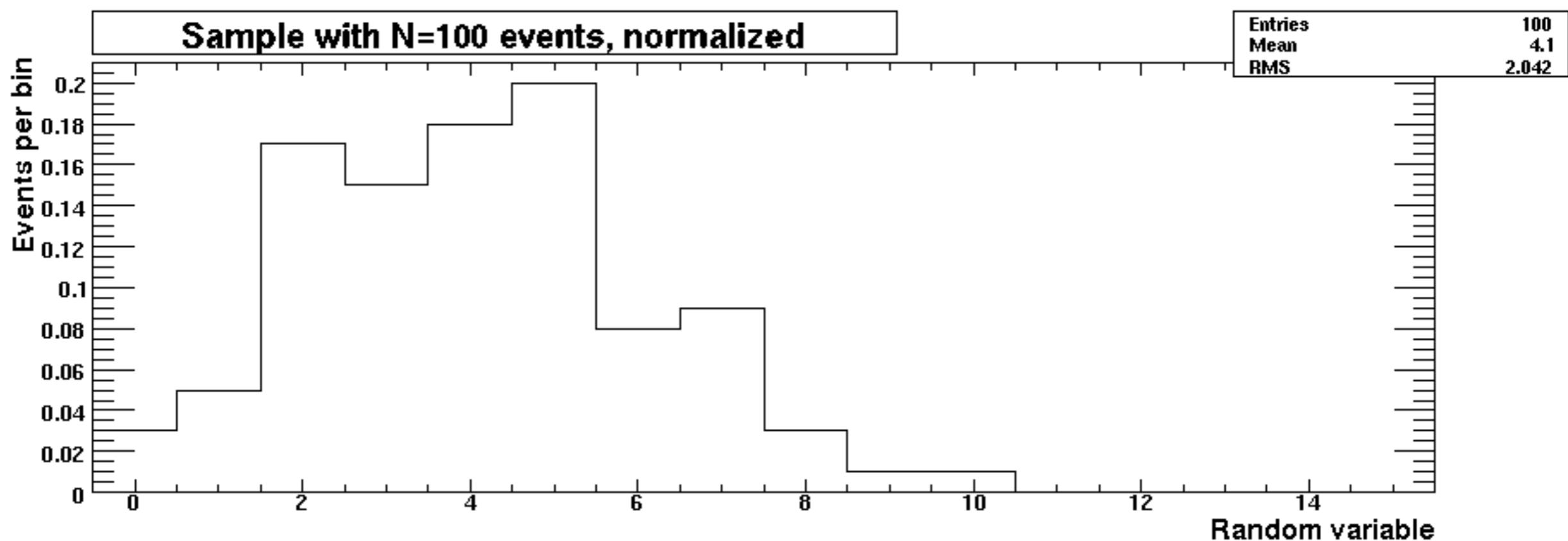
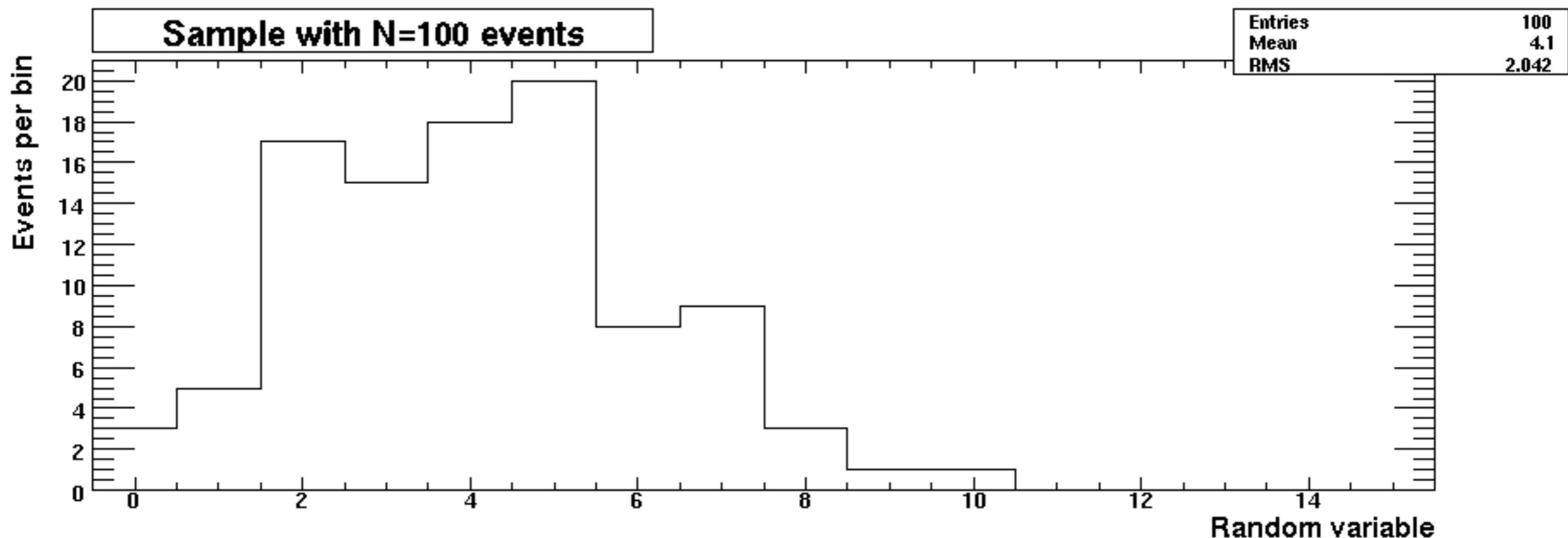
Sample From a Continuous PDF



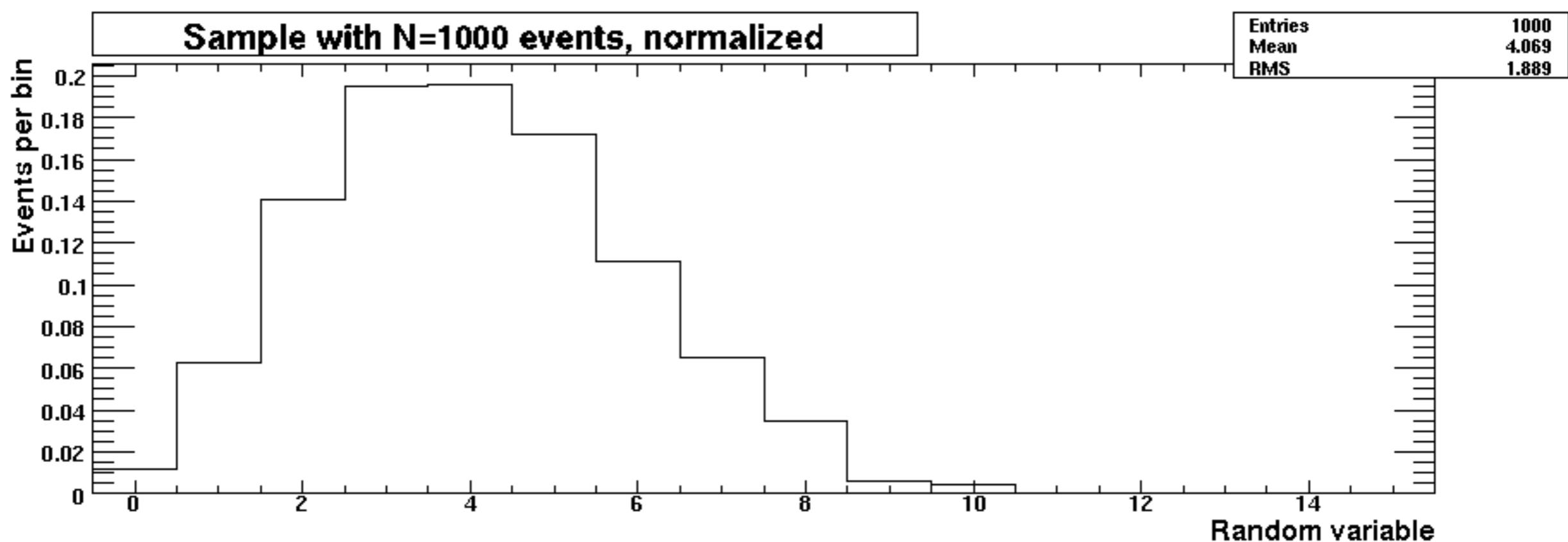
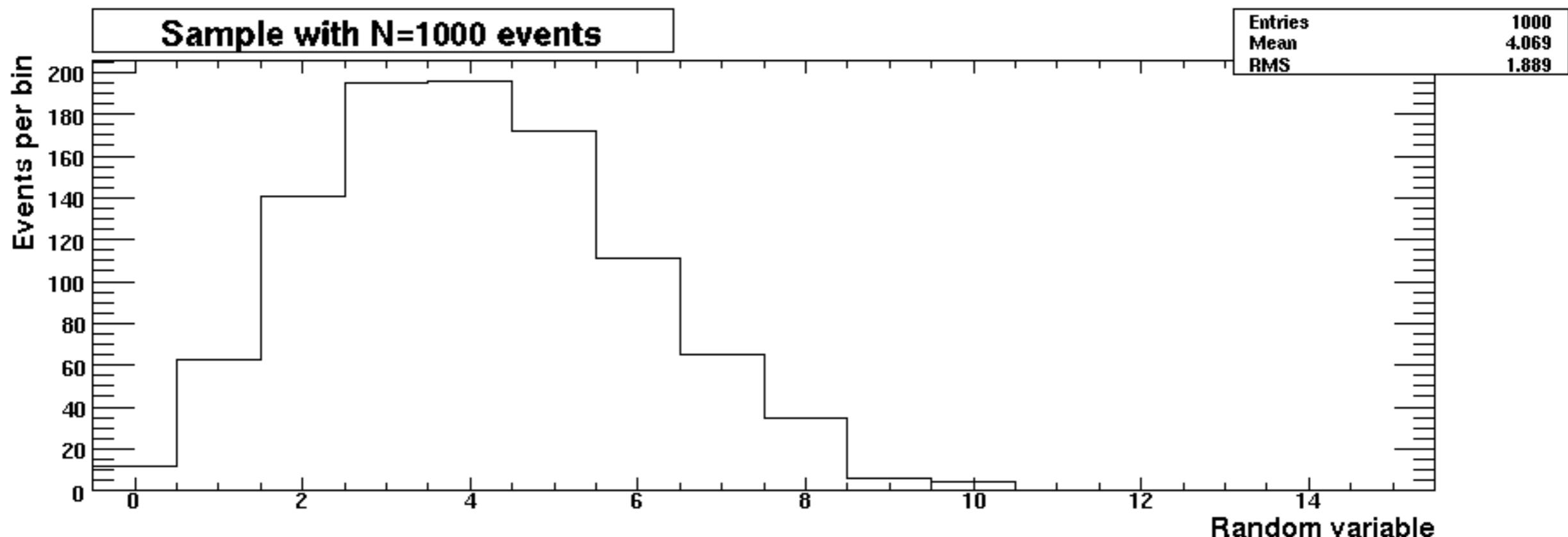
Sample From a Continuous PDF



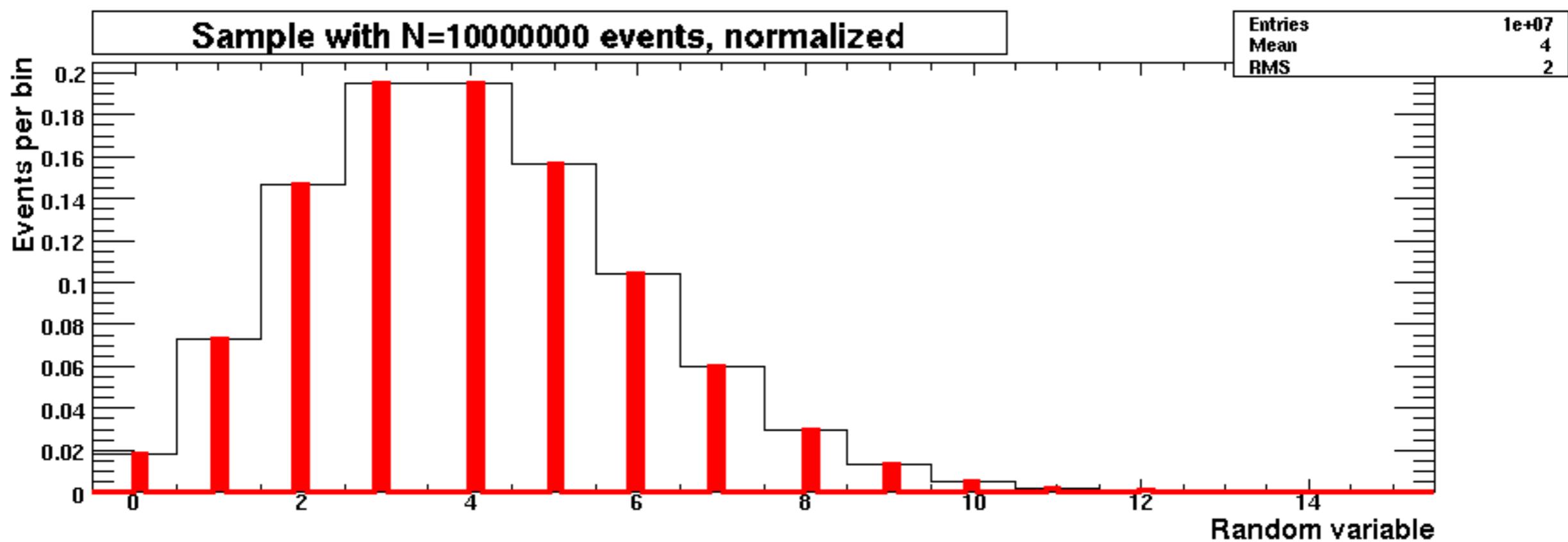
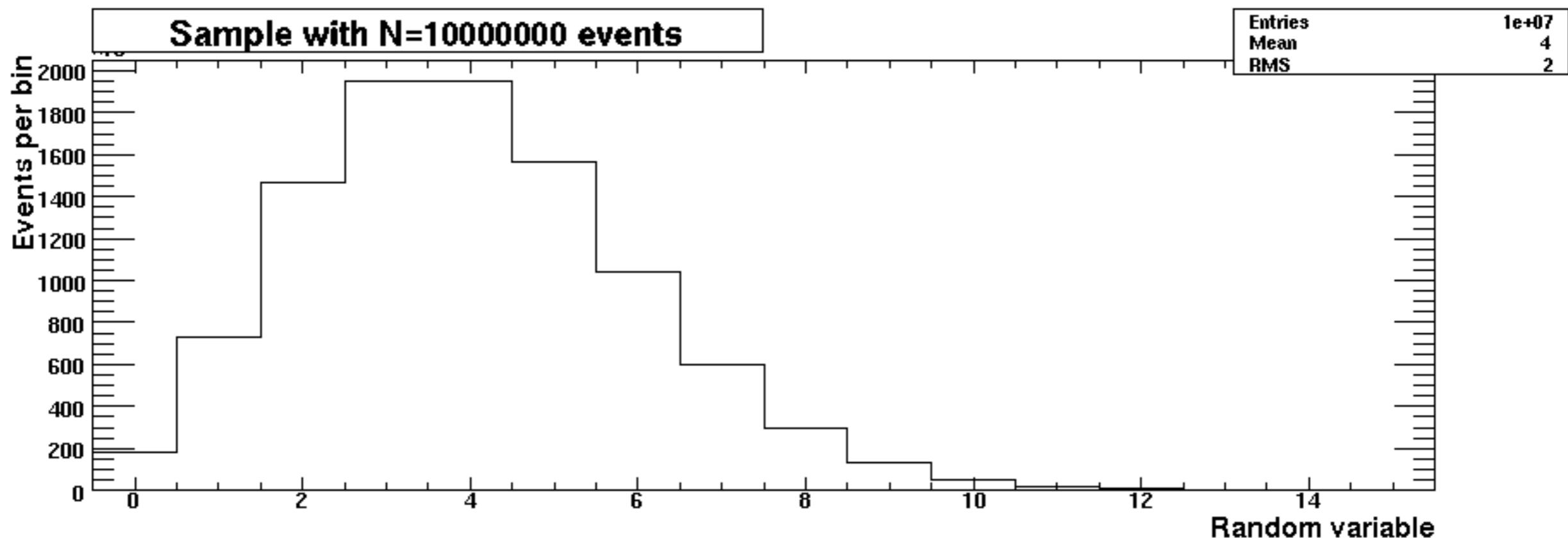
Sample From a Discrete PDF



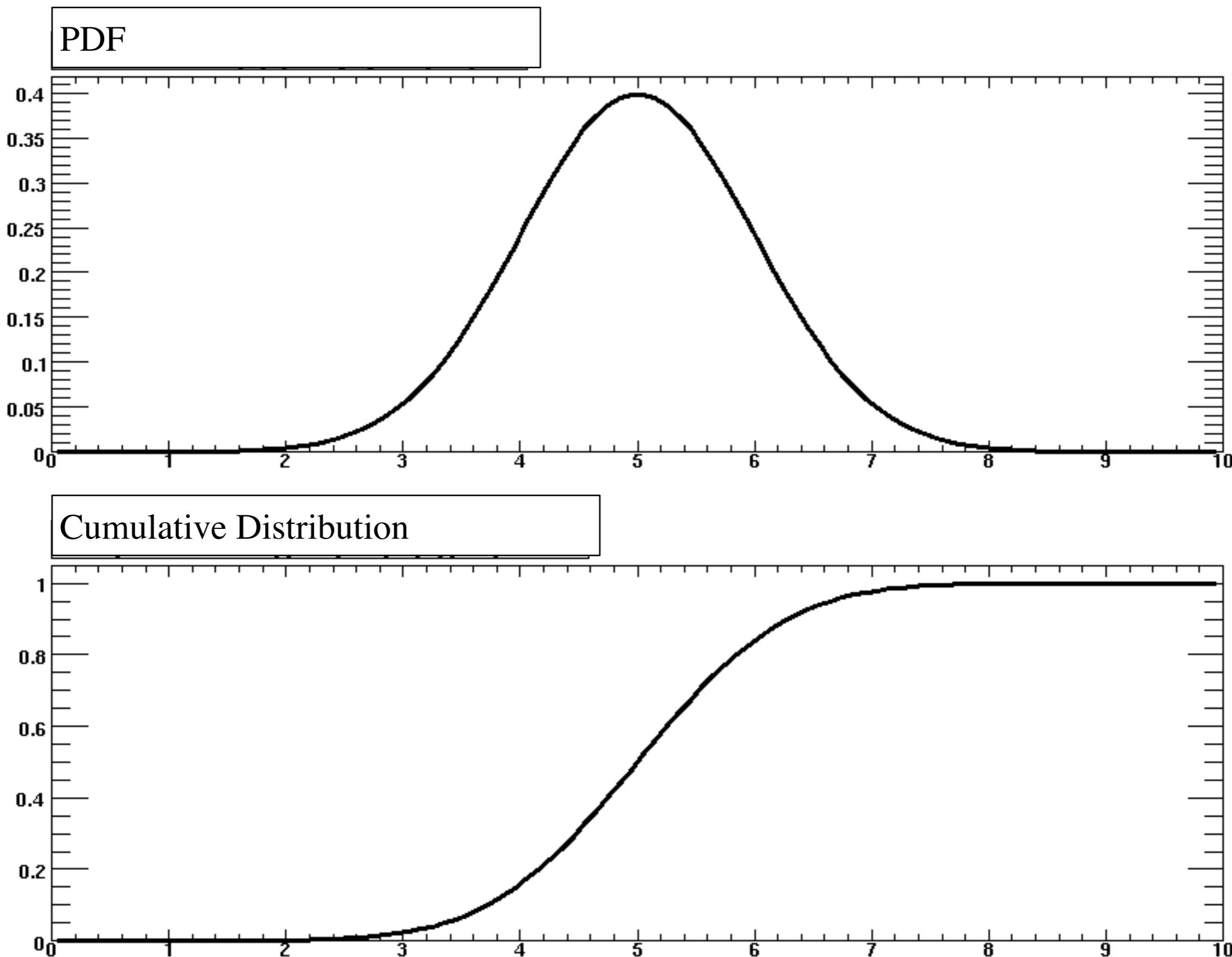
Sample From a Discrete PDF



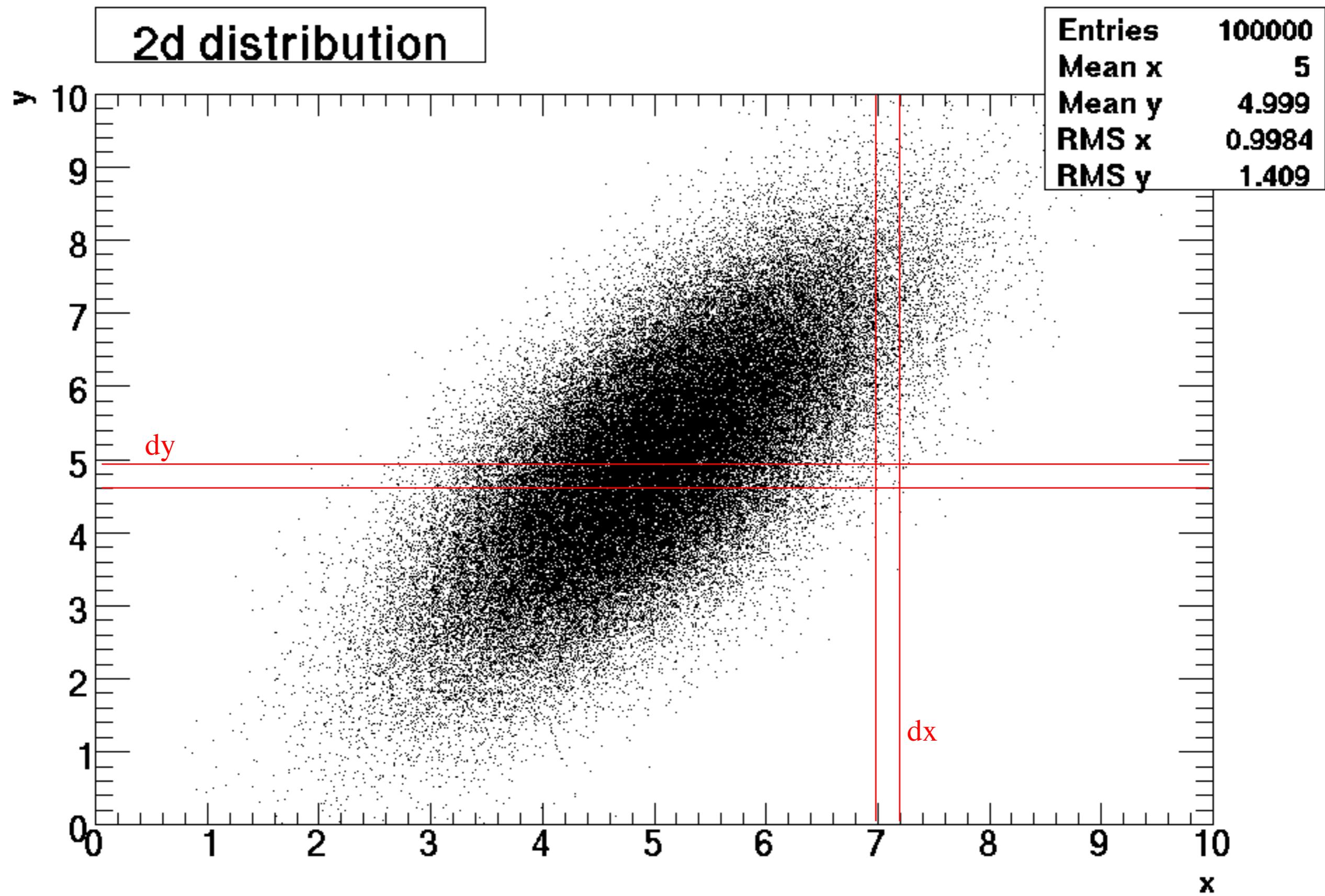
Sample From a Discrete PDF



Cumulative Distribution



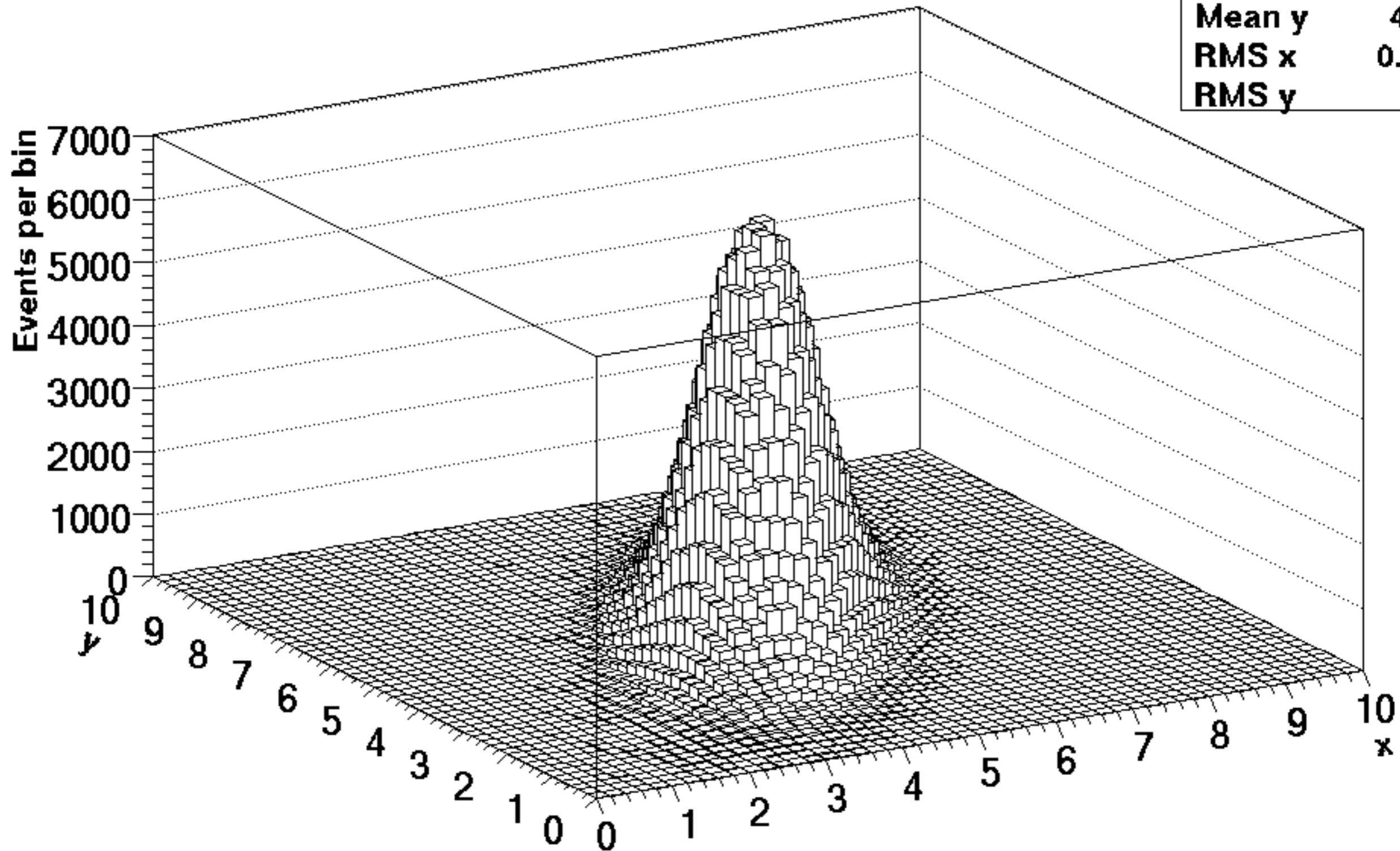
2d Distribution



2d Distribution

2d distribution

Entries	1000000
Mean x	4.999
Mean y	4.999
RMS x	0.9974
RMS y	1.41



Most important distributions

- or distribution
- distribution
- •
- distribution
- •
- There are many others
 -
 -
 -
 -

Binomial Distribution

- Random process with two possible outcomes

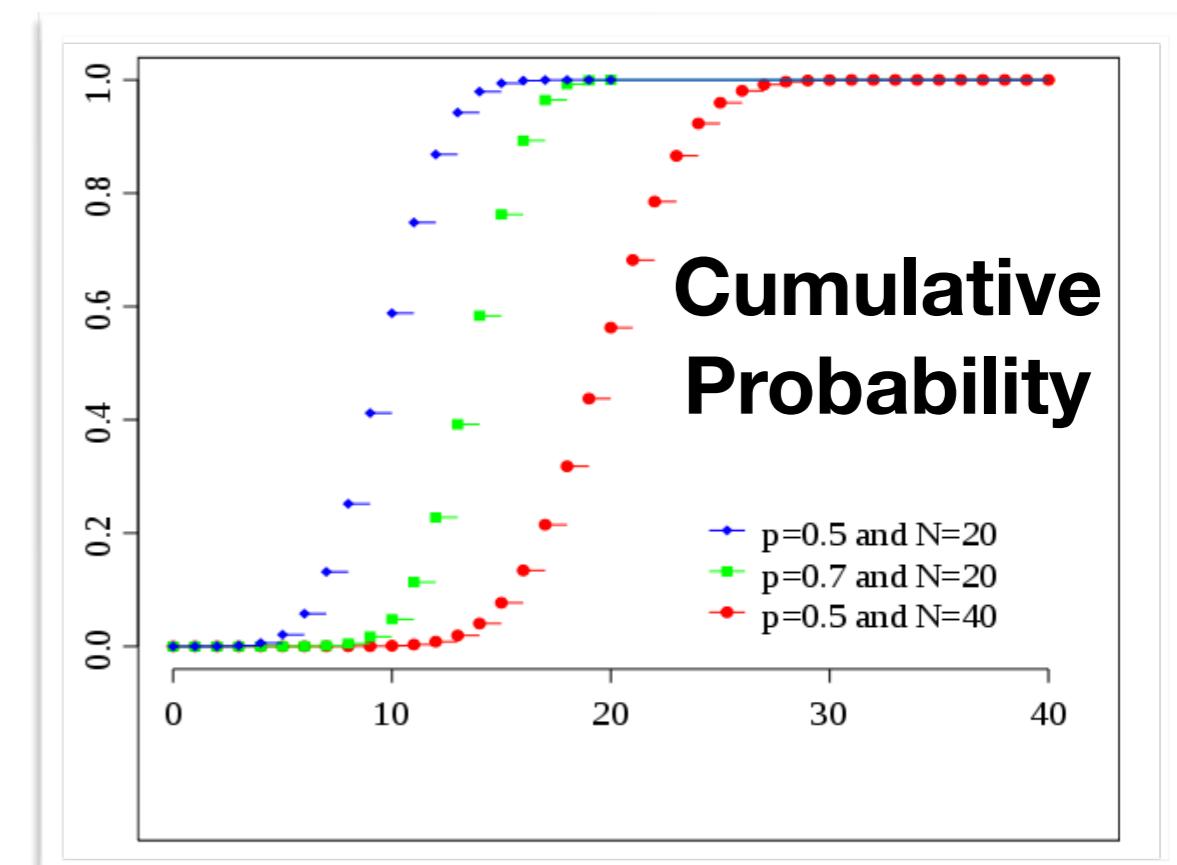
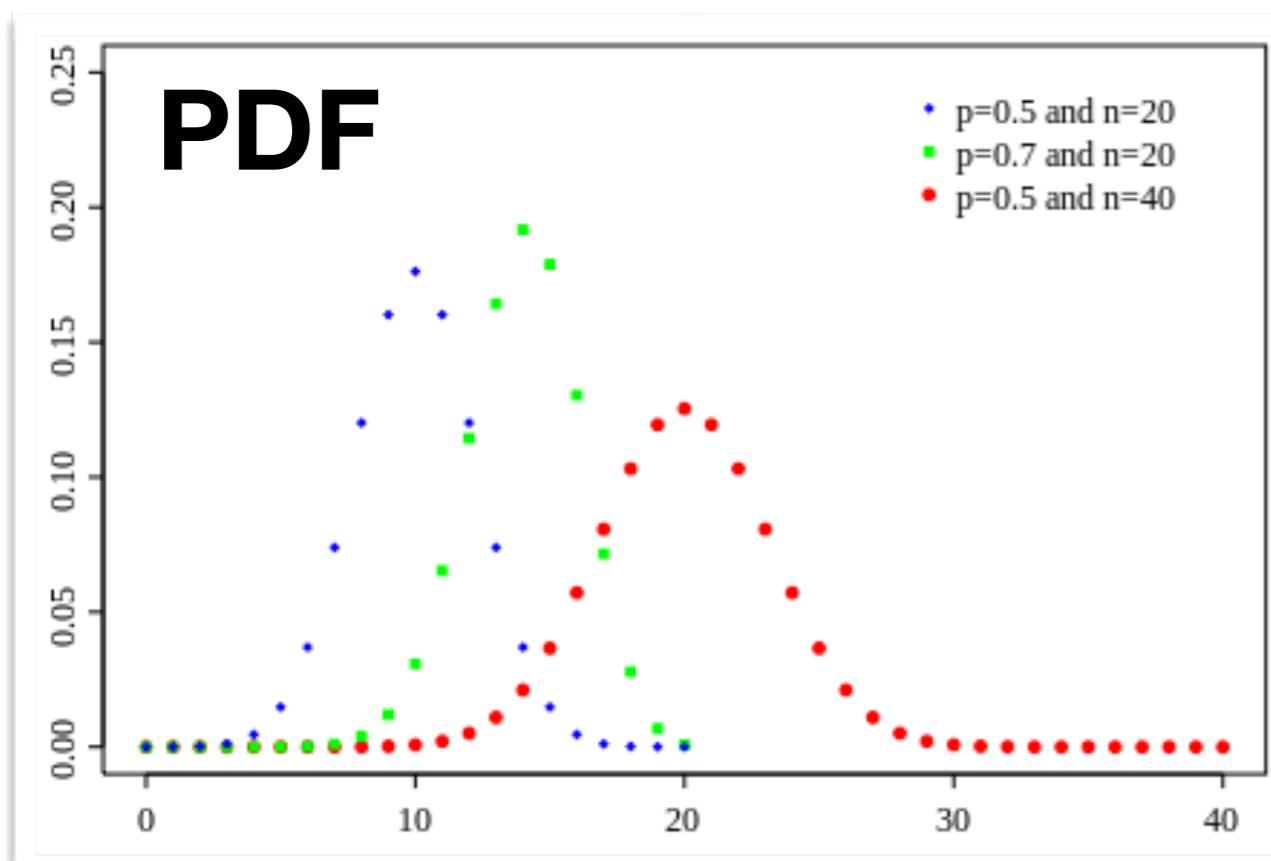
$$\bullet p = \quad q =$$

- After n trials, prob of getting outcome #1 exactly k times is

$$f(k, p) = \text{where } \binom{n}{k} =$$



$$\mu = \quad \sigma =$$

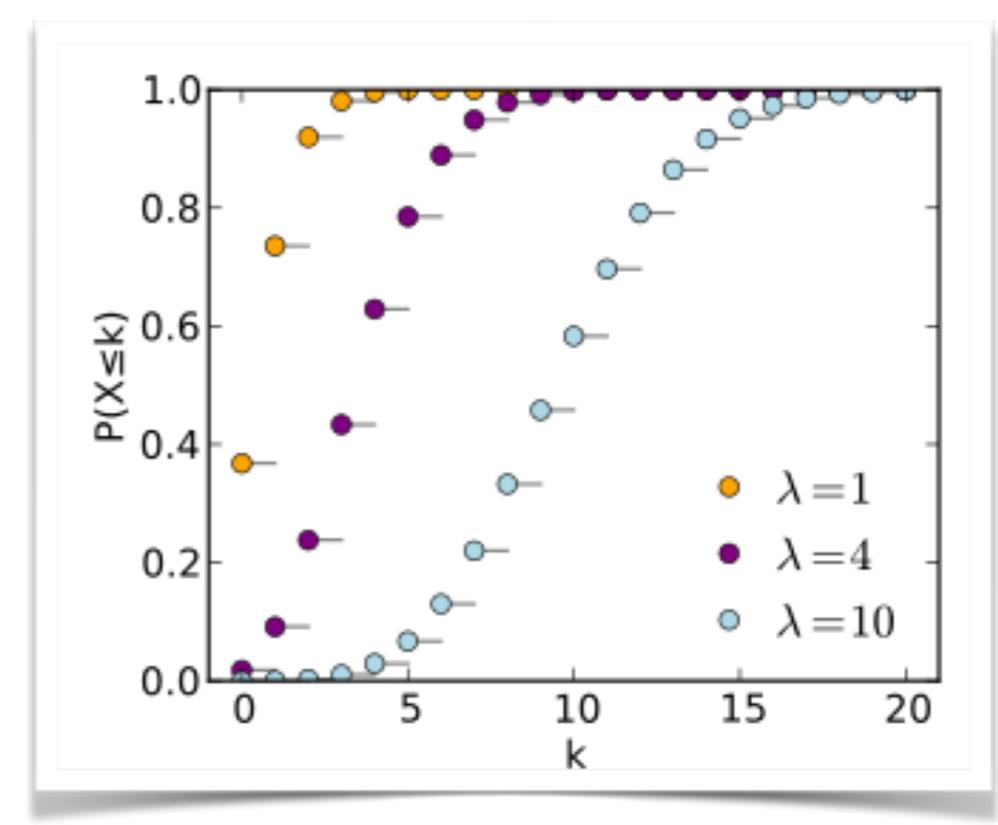
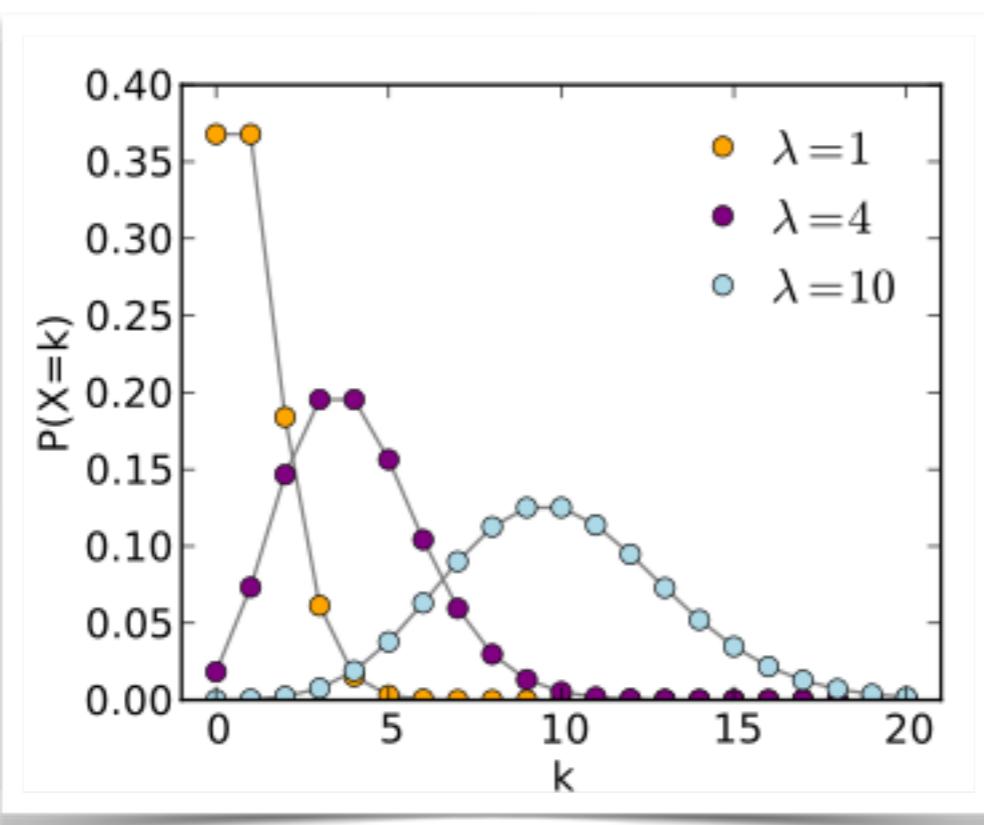


Poisson Distribution

- Prob of finding exactly k events in the interval if the events occur with an average rate λ
- $f(k; \lambda) =$
- For large λ approaches a Gaussian

$$\mu =$$

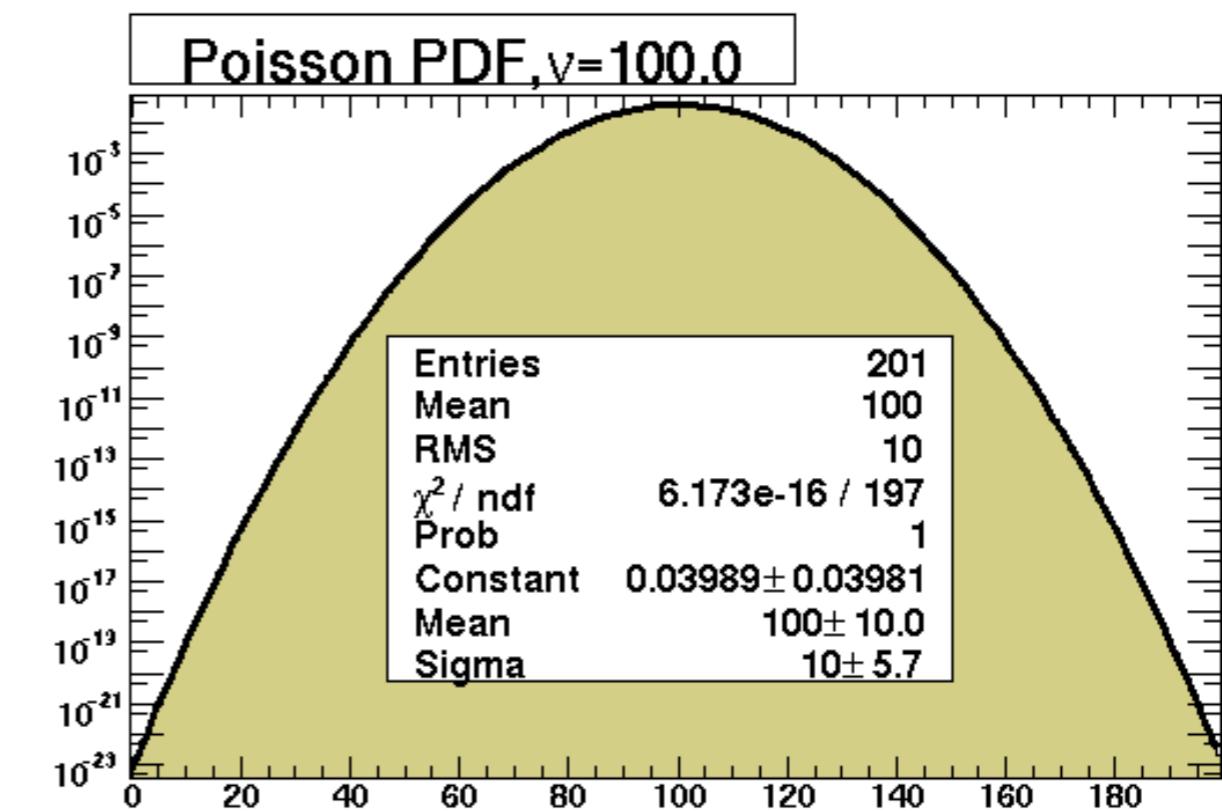
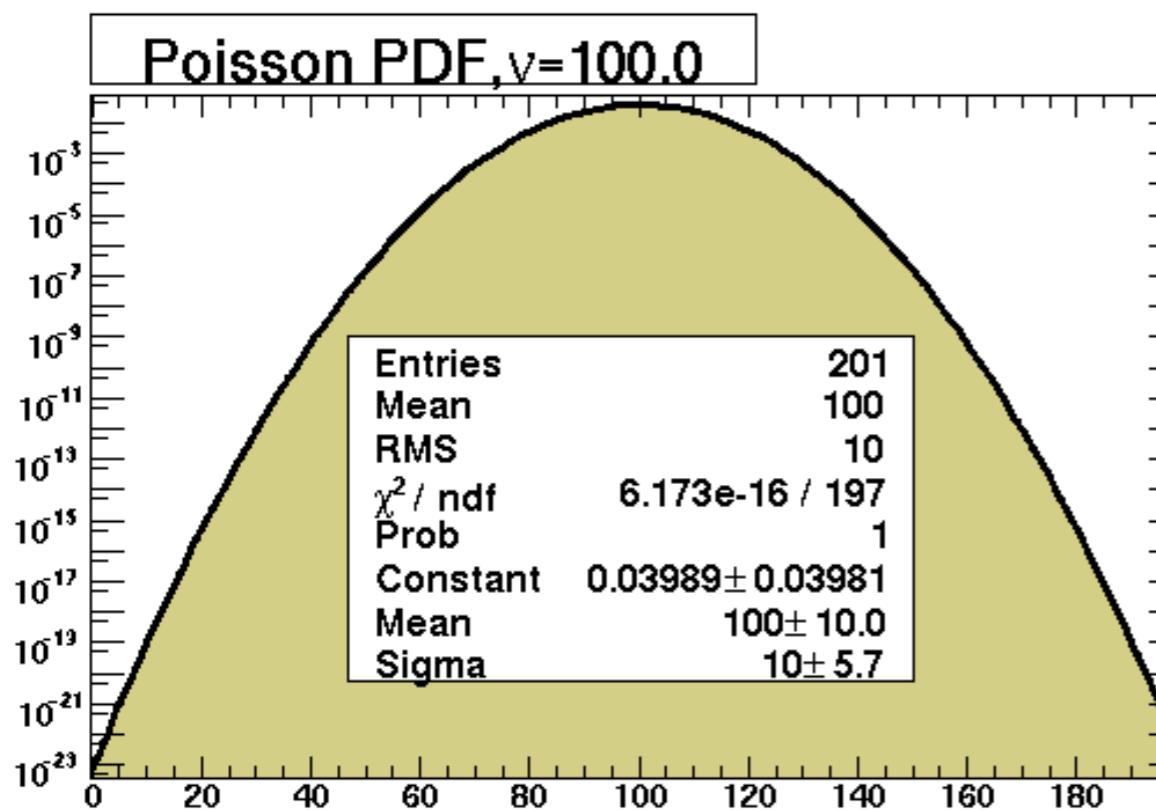
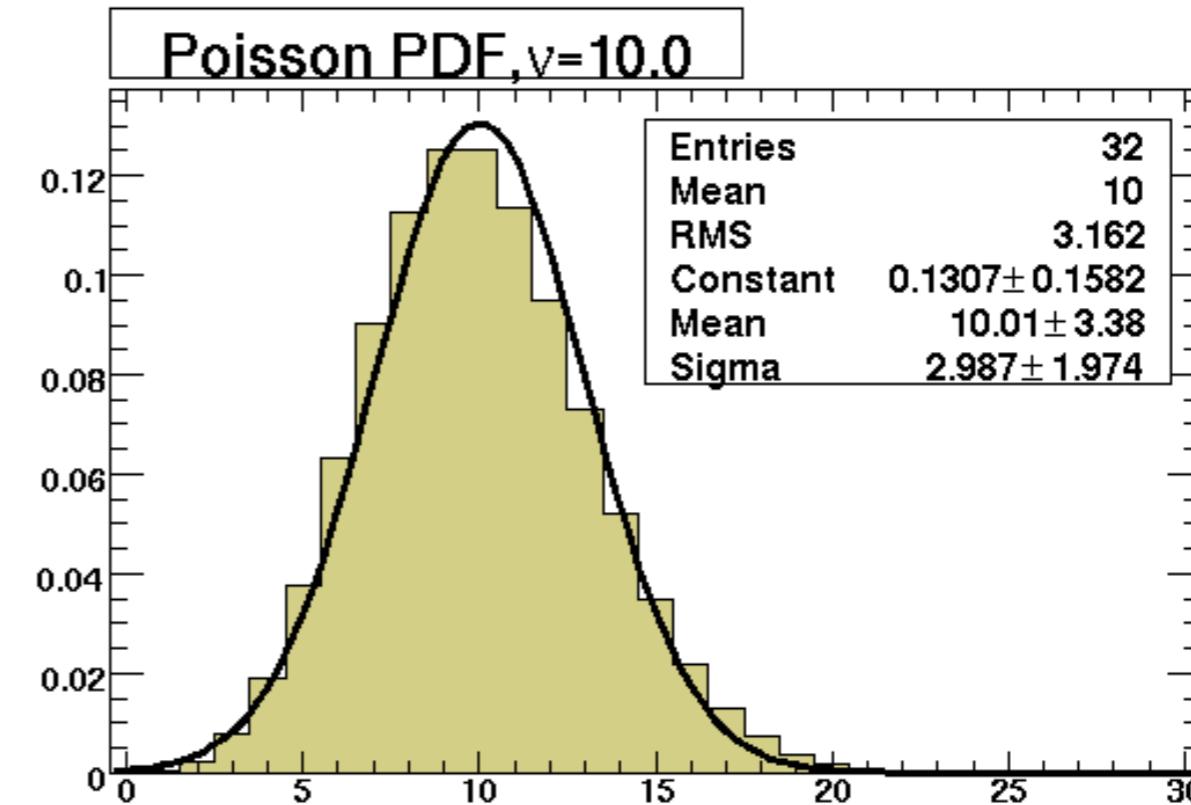
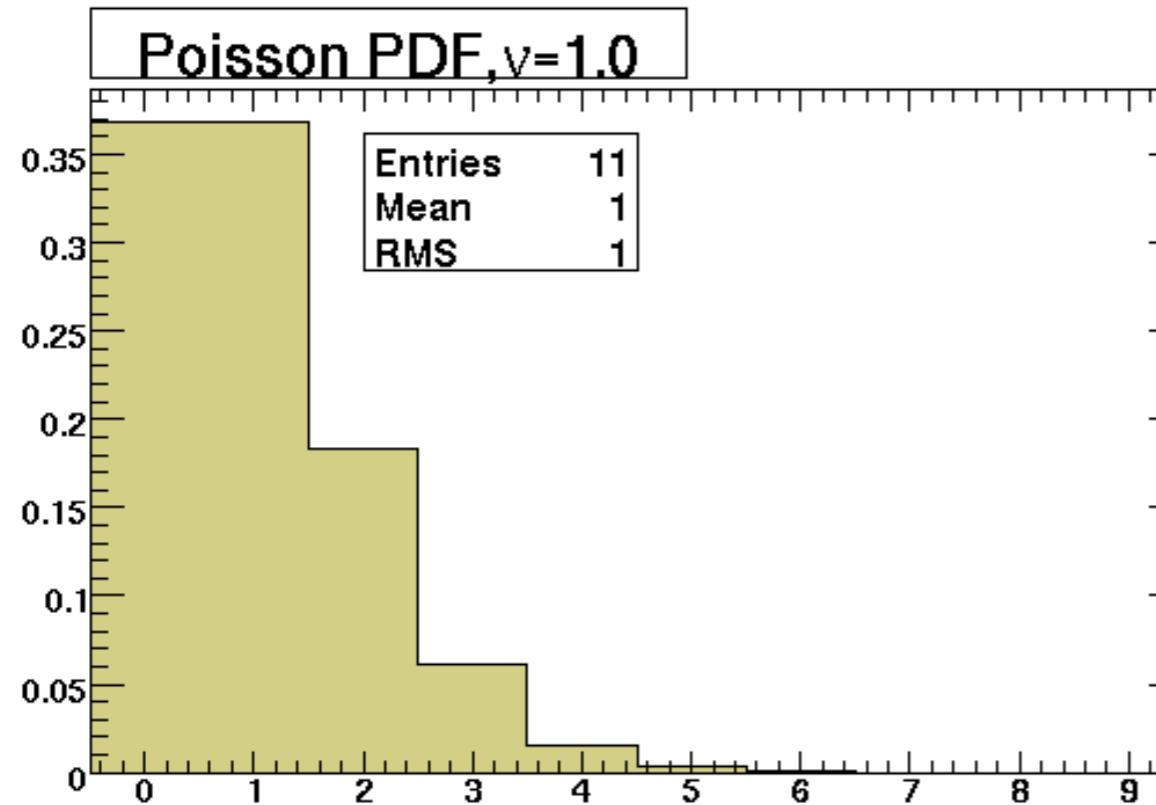
$$\sigma =$$



Example: Measure Efficiency

- Generate a sample of
- Apply selection; suppose events passed
- Estimate
 - $\hat{\epsilon} =$
 - $\sigma(\hat{\epsilon}) =$
 - $\sigma(\hat{\epsilon}) \neq$

Poisson Distribution



Gaussian Distribution

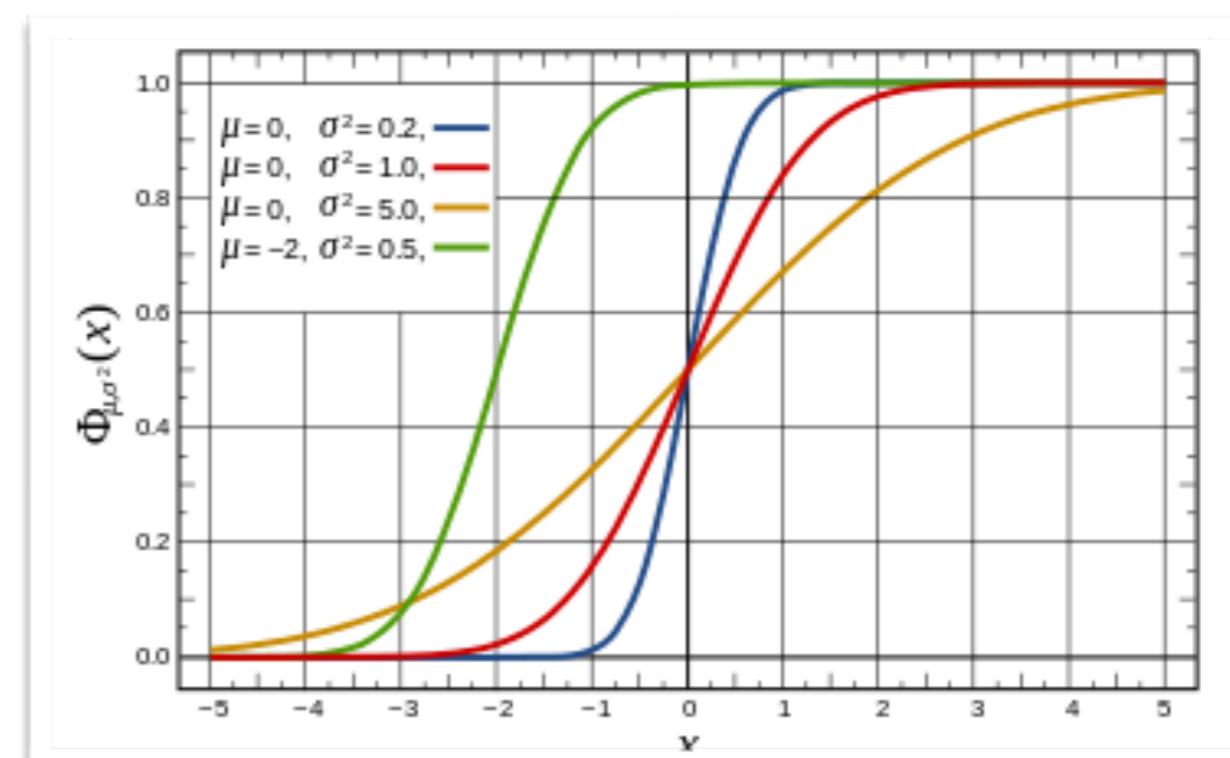
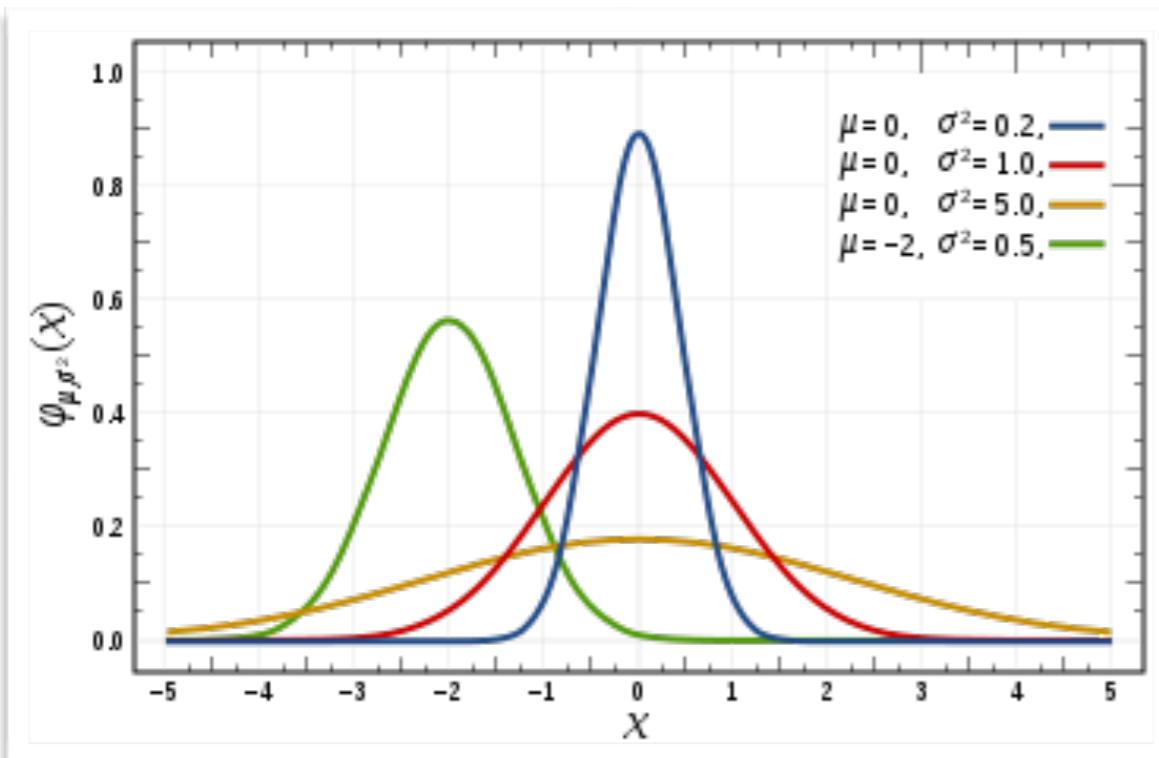
- **Central Limit Theorem**

- Given a random sample (x_1, x_2, \dots, x_n) drawn from a pdf with mean μ and variance σ^2 , if the mean is

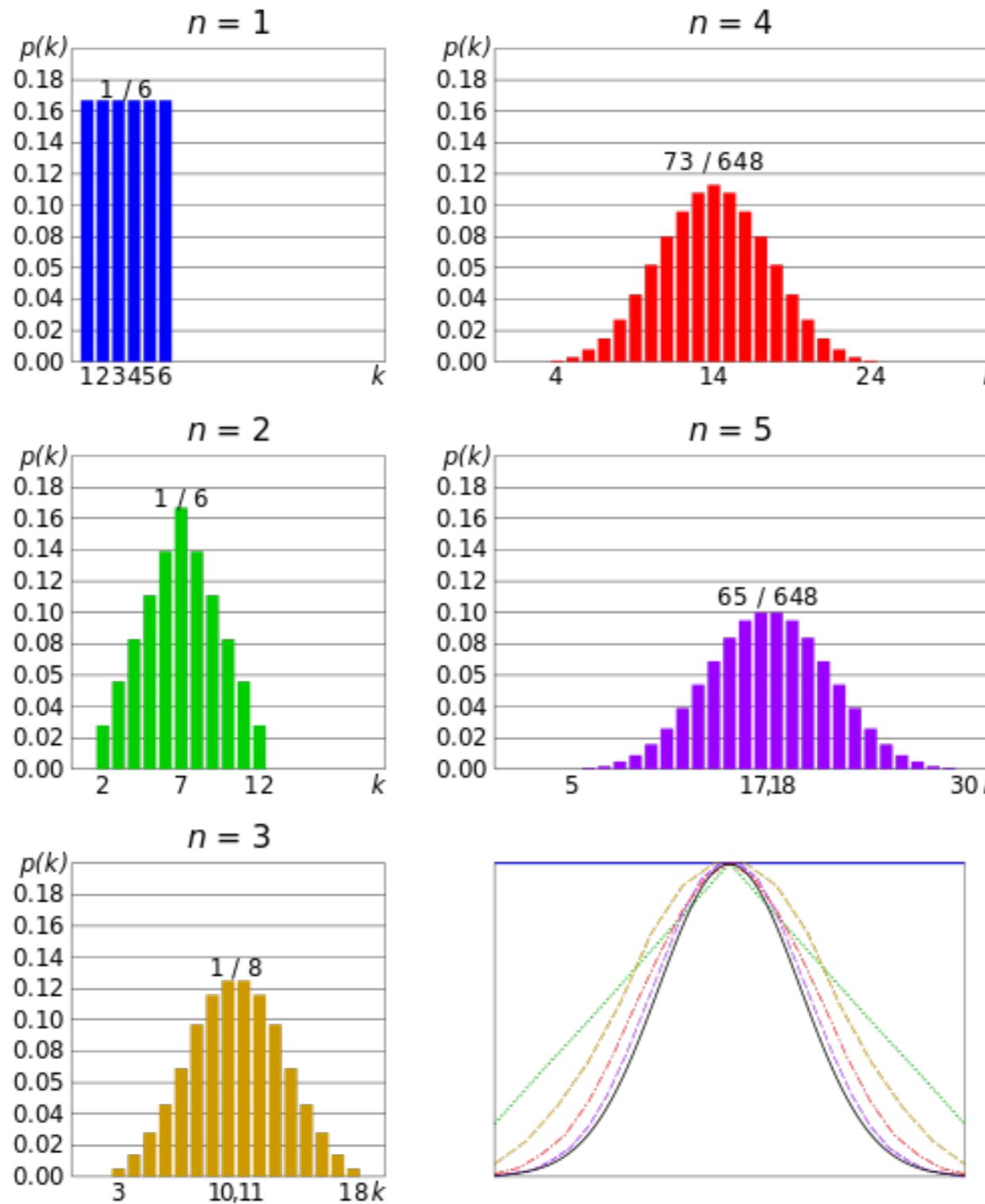
- $S/n =$

- the distribution of S/n approaches the distribution as

- $f(x; \mu, \sigma) =$



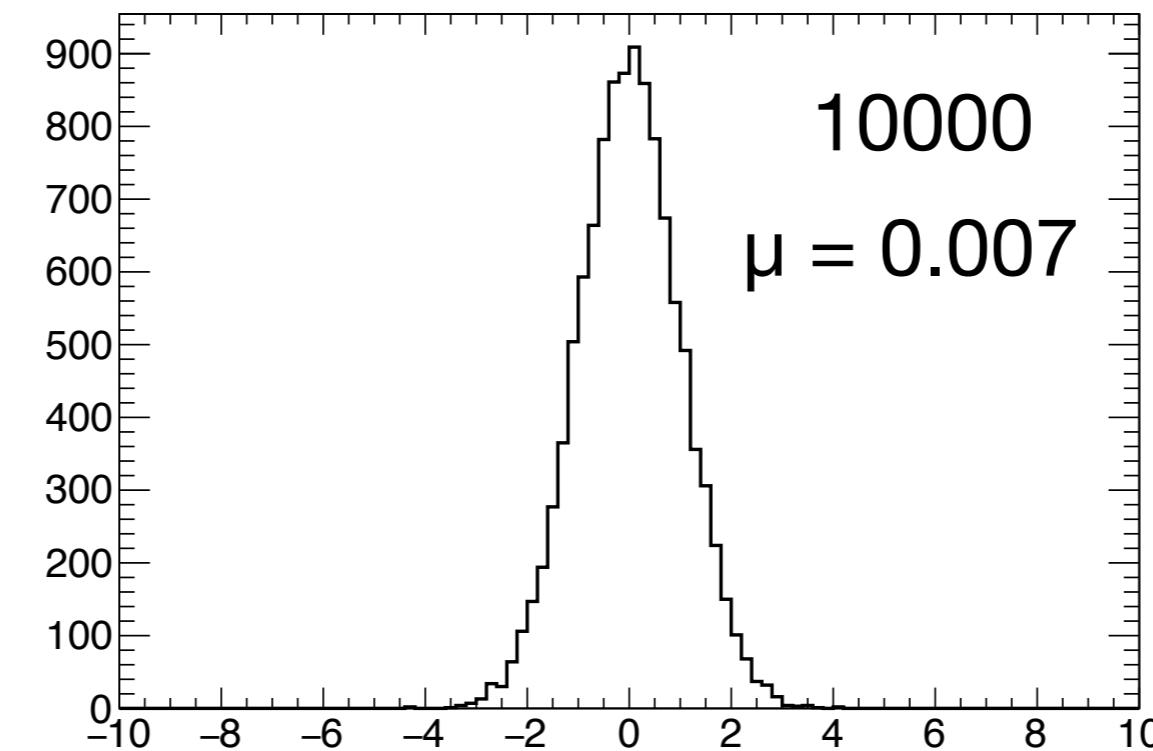
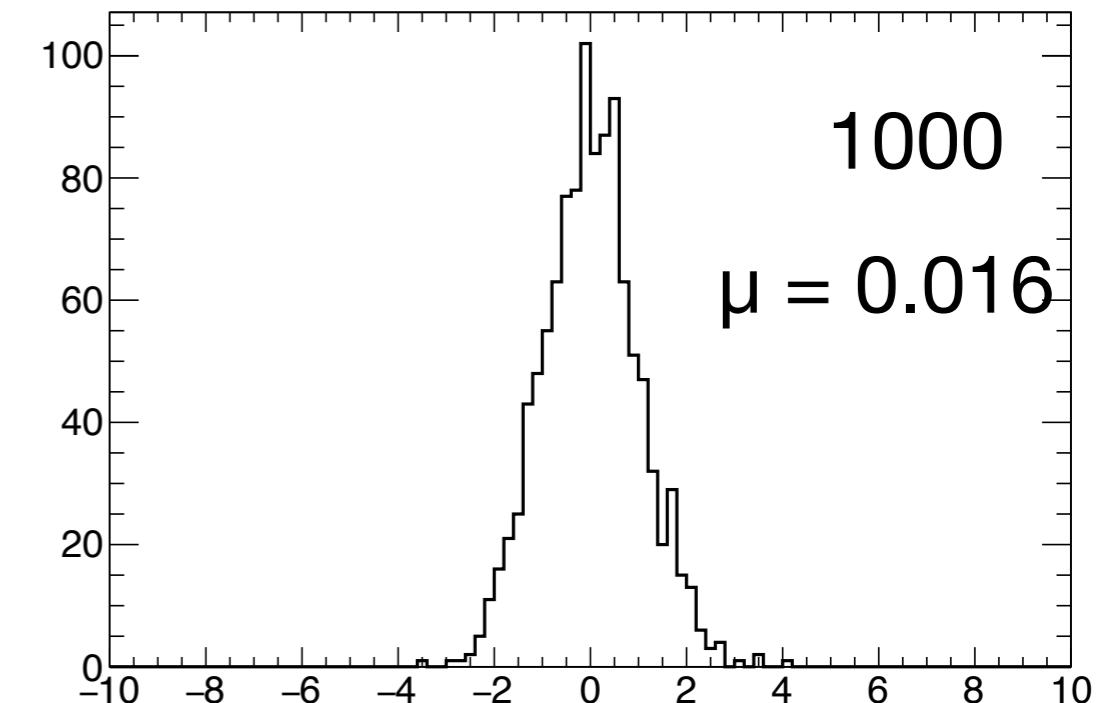
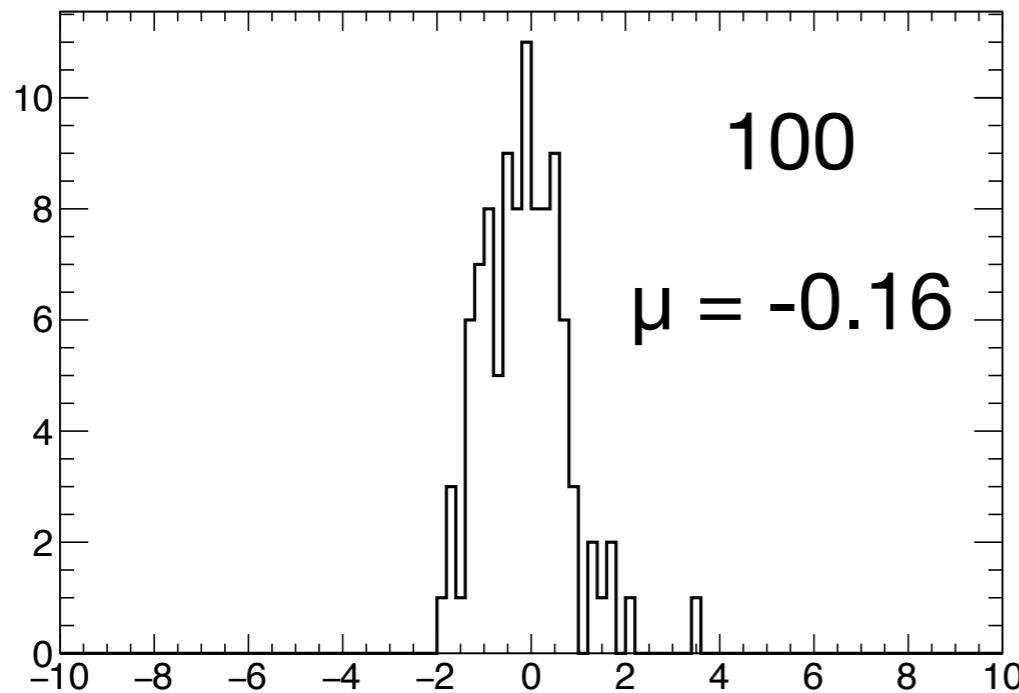
Central Limit Theorem



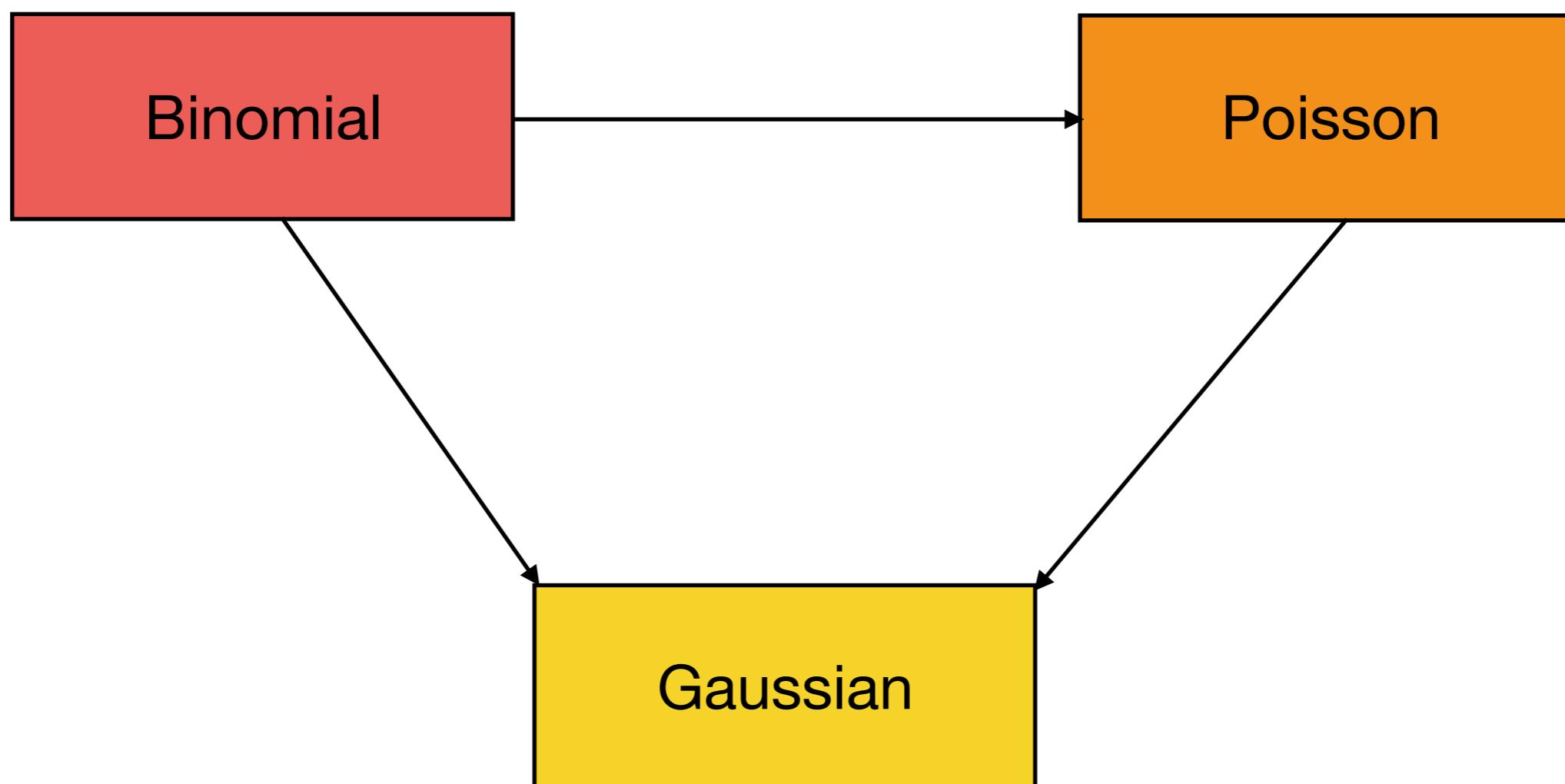
http://en.wikipedia.org/wiki/File:Dice_sum_central_limit_theorem.svg

Example

Example in jupyter notebook



Relationships between pdfs



Point Estimation

- Standard problem: set of described by
 - $f(x) \equiv$
- Point estimation:
 - $\hat{\theta} =$
 - Estimator of

Estimators

- Typical goal: estimate from experimental data and understand the uncertainty on that measurement
- **Characteristics** of an estimator
 -
 -
 -
 -
 -
- **Uncertainty**: how far the might be from our estimate due to statistical fluctuations in the

Basic Estimators

- Estimators for μ and σ^2
- Shape of the distribution
 - Most distributions are approximately normal, but may be skewed if μ or σ^2 is not readily available
- Convenient for linear data transformations
 - Convenient for linear functions of data
 - Convenient for linear functions of data
 - Automatic measure of spread
 - Be careful of outliers
 - (e.g. when $x_1 = 100$, $x_2 = 101$, $x_3 = 102$, $x_4 = 103$, $x_5 = 104$, $x_6 = 105$, $x_7 = 106$, $x_8 = 107$, $x_9 = 108$, $x_{10} = 109$, $x_{11} = 110$, $x_{12} = 111$, $x_{13} = 112$, $x_{14} = 113$, $x_{15} = 114$, $x_{16} = 115$, $x_{17} = 116$, $x_{18} = 117$, $x_{19} = 118$, $x_{20} = 119$, $x_{21} = 120$, $x_{22} = 121$, $x_{23} = 122$, $x_{24} = 123$, $x_{25} = 124$, $x_{26} = 125$, $x_{27} = 126$, $x_{28} = 127$, $x_{29} = 128$, $x_{30} = 129$, $x_{31} = 130$, $x_{32} = 131$, $x_{33} = 132$, $x_{34} = 133$, $x_{35} = 134$, $x_{36} = 135$, $x_{37} = 136$, $x_{38} = 137$, $x_{39} = 138$, $x_{40} = 139$, $x_{41} = 140$, $x_{42} = 141$, $x_{43} = 142$, $x_{44} = 143$, $x_{45} = 144$, $x_{46} = 145$, $x_{47} = 146$, $x_{48} = 147$, $x_{49} = 148$, $x_{50} = 149$, $x_{51} = 150$, $x_{52} = 151$, $x_{53} = 152$, $x_{54} = 153$, $x_{55} = 154$, $x_{56} = 155$, $x_{57} = 156$, $x_{58} = 157$, $x_{59} = 158$, $x_{60} = 159$, $x_{61} = 160$, $x_{62} = 161$, $x_{63} = 162$, $x_{64} = 163$, $x_{65} = 164$, $x_{66} = 165$, $x_{67} = 166$, $x_{68} = 167$, $x_{69} = 168$, $x_{70} = 169$, $x_{71} = 170$, $x_{72} = 171$, $x_{73} = 172$, $x_{74} = 173$, $x_{75} = 174$, $x_{76} = 175$, $x_{77} = 176$, $x_{78} = 177$, $x_{79} = 178$, $x_{80} = 179$, $x_{81} = 180$, $x_{82} = 181$, $x_{83} = 182$, $x_{84} = 183$, $x_{85} = 184$, $x_{86} = 185$, $x_{87} = 186$, $x_{88} = 187$, $x_{89} = 188$, $x_{90} = 189$, $x_{91} = 190$, $x_{92} = 191$, $x_{93} = 192$, $x_{94} = 193$, $x_{95} = 194$, $x_{96} = 195$, $x_{97} = 196$, $x_{98} = 197$, $x_{99} = 198$, $x_{100} = 199$, $x_{101} = 200$, $x_{102} = 201$, $x_{103} = 202$, $x_{104} = 203$, $x_{105} = 204$, $x_{106} = 205$, $x_{107} = 206$, $x_{108} = 207$, $x_{109} = 208$, $x_{110} = 209$, $x_{111} = 210$, $x_{112} = 211$, $x_{113} = 212$, $x_{114} = 213$, $x_{115} = 214$, $x_{116} = 215$, $x_{117} = 216$, $x_{118} = 217$, $x_{119} = 218$, $x_{120} = 219$, $x_{121} = 220$, $x_{122} = 221$, $x_{123} = 222$, $x_{124} = 223$, $x_{125} = 224$, $x_{126} = 225$, $x_{127} = 226$, $x_{128} = 227$, $x_{129} = 228$, $x_{130} = 229$, $x_{131} = 230$, $x_{132} = 231$, $x_{133} = 232$, $x_{134} = 233$, $x_{135} = 234$, $x_{136} = 235$, $x_{137} = 236$, $x_{138} = 237$, $x_{139} = 238$, $x_{140} = 239$, $x_{141} = 240$, $x_{142} = 241$, $x_{143} = 242$, $x_{144} = 243$, $x_{145} = 244$, $x_{146} = 245$, $x_{147} = 246$, $x_{148} = 247$, $x_{149} = 248$, $x_{150} = 249$, $x_{151} = 250$, $x_{152} = 251$, $x_{153} = 252$, $x_{154} = 253$, $x_{155} = 254$, $x_{156} = 255$, $x_{157} = 256$, $x_{158} = 257$, $x_{159} = 258$, $x_{160} = 259$, $x_{161} = 260$, $x_{162} = 261$, $x_{163} = 262$, $x_{164} = 263$, $x_{165} = 264$, $x_{166} = 265$, $x_{167} = 266$, $x_{168} = 267$, $x_{169} = 268$, $x_{170} = 269$, $x_{171} = 270$, $x_{172} = 271$, $x_{173} = 272$, $x_{174} = 273$, $x_{175} = 274$, $x_{176} = 275$, $x_{177} = 276$, $x_{178} = 277$, $x_{179} = 278$, $x_{180} = 279$, $x_{181} = 280$, $x_{182} = 281$, $x_{183} = 282$, $x_{184} = 283$, $x_{185} = 284$, $x_{186} = 285$, $x_{187} = 286$, $x_{188} = 287$, $x_{189} = 288$, $x_{190} = 289$, $x_{191} = 290$, $x_{192} = 291$, $x_{193} = 292$, $x_{194} = 293$, $x_{195} = 294$, $x_{196} = 295$, $x_{197} = 296$, $x_{198} = 297$, $x_{199} = 298$, $x_{200} = 299$, $x_{201} = 300$, $x_{202} = 301$, $x_{203} = 302$, $x_{204} = 303$, $x_{205} = 304$, $x_{206} = 305$, $x_{207} = 306$, $x_{208} = 307$, $x_{209} = 308$, $x_{210} = 309$, $x_{211} = 310$, $x_{212} = 311$, $x_{213} = 312$, $x_{214} = 313$, $x_{215} = 314$, $x_{216} = 315$, $x_{217} = 316$, $x_{218} = 317$, $x_{219} = 318$, $x_{220} = 319$, $x_{221} = 320$, $x_{222} = 321$, $x_{223} = 322$, $x_{224} = 323$, $x_{225} = 324$, $x_{226} = 325$, $x_{227} = 326$, $x_{228} = 327$, $x_{229} = 328$, $x_{230} = 329$, $x_{231} = 330$, $x_{232} = 331$, $x_{233} = 332$, $x_{234} = 333$, $x_{235} = 334$, $x_{236} = 335$, $x_{237} = 336$, $x_{238} = 337$, $x_{239} = 338$, $x_{240} = 339$, $x_{241} = 340$, $x_{242} = 341$, $x_{243} = 342$, $x_{244} = 343$, $x_{245} = 344$, $x_{246} = 345$, $x_{247} = 346$, $x_{248} = 347$, $x_{249} = 348$, $x_{250} = 349$, $x_{251} = 350$, $x_{252} = 351$, $x_{253} = 352$, $x_{254} = 353$, $x_{255} = 354$, $x_{256} = 355$, $x_{257} = 356$, $x_{258} = 357$, $x_{259} = 358$, $x_{260} = 359$, $x_{261} = 360$, $x_{262} = 361$, $x_{263} = 362$, $x_{264} = 363$, $x_{265} = 364$, $x_{266} = 365$, $x_{267} = 366$, $x_{268} = 367$, $x_{269} = 368$, $x_{270} = 369$, $x_{271} = 370$, $x_{272} = 371$, $x_{273} = 372$, $x_{274} = 373$, $x_{275} = 374$, $x_{276} = 375$, $x_{277} = 376$, $x_{278} = 377$, $x_{279} = 378$, $x_{280} = 379$, $x_{281} = 380$, $x_{282} = 381$, $x_{283} = 382$, $x_{284} = 383$, $x_{285} = 384$, $x_{286} = 385$, $x_{287} = 386$, $x_{288} = 387$, $x_{289} = 388$, $x_{290} = 389$, $x_{291} = 390$, $x_{292} = 391$, $x_{293} = 392$, $x_{294} = 393$, $x_{295} = 394$, $x_{296} = 395$, $x_{297} = 396$, $x_{298} = 397$, $x_{299} = 398$, $x_{300} = 399$, $x_{301} = 400$, $x_{302} = 401$, $x_{303} = 402$, $x_{304} = 403$, $x_{305} = 404$, $x_{306} = 405$, $x_{307} = 406$, $x_{308} = 407$, $x_{309} = 408$, $x_{310} = 409$, $x_{311} = 410$, $x_{312} = 411$, $x_{313} = 412$, $x_{314} = 413$, $x_{315} = 414$, $x_{316} = 415$, $x_{317} = 416$, $x_{318} = 417$, $x_{319} = 418$, $x_{320} = 419$, $x_{321} = 420$, $x_{322} = 421$, $x_{323} = 422$, $x_{324} = 423$, $x_{325} = 424$, $x_{326} = 425$, $x_{327} = 426$, $x_{328} = 427$, $x_{329} = 428$, $x_{330} = 429$, $x_{331} = 430$, $x_{332} = 431$, $x_{333} = 432$, $x_{334} = 433$, $x_{335} = 434$, $x_{336} = 435$, $x_{337} = 436$, $x_{338} = 437$, $x_{339} = 438$, $x_{340} = 439$, $x_{341} = 440$, $x_{342} = 441$, $x_{343} = 442$, $x_{344} = 443$, $x_{345} = 444$, $x_{346} = 445$, $x_{347} = 446$, $x_{348} = 447$, $x_{349} = 448$, $x_{350} = 449$, $x_{351} = 450$, $x_{352} = 451$, $x_{353} = 452$, $x_{354} = 453$, $x_{355} = 454$, $x_{356} = 455$, $x_{357} = 456$, $x_{358} = 457$, $x_{359} = 458$, $x_{360} = 459$, $x_{361} = 460$, $x_{362} = 461$, $x_{363} = 462$, $x_{364} = 463$, $x_{365} = 464$, $x_{366} = 465$, $x_{367} = 466$, $x_{368} = 467$, $x_{369} = 468$, $x_{370} = 469$, $x_{371} = 470$, $x_{372} = 471$, $x_{373} = 472$, $x_{374} = 473$, $x_{375} = 474$, $x_{376} = 475$, $x_{377} = 476$, $x_{378} = 477$, $x_{379} = 478$, $x_{380} = 479$, $x_{381} = 480$, $x_{382} = 481$, $x_{383} = 482$, $x_{384} = 483$, $x_{385} = 484$, $x_{386} = 485$, $x_{387} = 486$, $x_{388} = 487$, $x_{389} = 488$, $x_{390} = 489$, $x_{391} = 490$, $x_{392} = 491$, $x_{393} = 492$, $x_{394} = 493$, $x_{395} = 494$, $x_{396} = 495$, $x_{397} = 496$, $x_{398} = 497$, $x_{399} = 498$, $x_{400} = 499$, $x_{401} = 500$, $x_{402} = 501$, $x_{403} = 502$, $x_{404} = 503$, $x_{405} = 504$, $x_{406} = 505$, $x_{407} = 506$, $x_{408} = 507$, $x_{409} = 508$, $x_{410} = 509$, $x_{411} = 510$, $x_{412} = 511$, $x_{413} = 512$, $x_{414} = 513$, $x_{415} = 514$, $x_{416} = 515$, $x_{417} = 516$, $x_{418} = 517$, $x_{419} = 518$, $x_{420} = 519$, $x_{421} = 520$, $x_{422} = 521$, $x_{423} = 522$, $x_{424} = 523$, $x_{425} = 524$, $x_{426} = 525$, $x_{427} = 526$, $x_{428} = 527$, $x_{429} = 528$, $x_{430} = 529$, $x_{431} = 530$, $x_{432} = 531$, $x_{433} = 532$, $x_{434} = 533$, $x_{435} = 534$, $x_{436} = 535$, $x_{437} = 536$, $x_{438} = 537$, $x_{439} = 538$, $x_{440} = 539$, $x_{441} = 540$, $x_{442} = 541$, $x_{443} = 542$, $x_{444} = 543$, $x_{445} = 544$, $x_{446} = 545$, $x_{447} = 546$, $x_{448} = 547$, $x_{449} = 548$, $x_{450} = 549$, $x_{451} = 550$, $x_{452} = 551$, $x_{453} = 552$, $x_{454} = 553$, $x_{455} = 554$, $x_{456} = 555$, $x_{457} = 556$, $x_{458} = 557$, $x_{459} = 558$, $x_{460} = 559$, $x_{461} = 560$, $x_{462} = 561$, $x_{463} = 562$, $x_{464} = 563$, $x_{465} = 564$, $x_{466} = 565$, $x_{467} = 566$, $x_{468} = 567$, $x_{469} = 568$, $x_{470} = 569$, $x_{471} = 570$, $x_{472} = 571$, $x_{473} = 572$, $x_{474} = 573$, $x_{475} = 574$, $x_{476} = 575$, $x_{477} = 576$, $x_{478} = 577$, $x_{479} = 578$, $x_{480} = 579$, $x_{481} = 580$, $x_{482} = 581$, $x_{483} = 582$, $x_{484} = 583$, $x_{485} = 584$, $x_{486} = 585$, $x_{487} = 586$, $x_{488} = 587$, $x_{489} = 588$, $x_{490} = 589$, $x_{491} = 590$, $x_{492} = 591$, $x_{493} = 592$, $x_{494} = 593$, $x_{495} = 594$, $x_{496} = 595$, $x_{497} = 596$, $x_{498} = 597$, $x_{499} = 598$, $x_{500} = 599$, $x_{501} = 600$, $x_{502} = 601$, $x_{503} = 602$, $x_{504} = 603$, $x_{505} = 604$, $x_{506} = 605$, $x_{507} = 606$, $x_{508} = 607$, $x_{509} = 608$, $x_{510} = 609$, $x_{511} = 610$, $x_{512} = 611$, $x_{513} = 612$, $x_{514} = 613$, $x_{515} = 614$, $x_{516} = 615$, $x_{517} = 616$, $x_{518} = 617$, $x_{519} = 618$, $x_{520} = 619$, $x_{521} = 620$, $x_{522} = 621$, $x_{523} = 622$, $x_{524} = 623$, $x_{525} = 624$, $x_{526} = 625$, $x_{527} = 626$, $x_{528} = 627$, $x_{529} = 628$, $x_{530} =$

Mean and Variance from a Sample

- Estimators (equally data)

- $\hat{\mu} =$

- $\hat{\sigma}^2 =$

- Variances of these

- $V[\hat{\mu}] =$

- $V[\hat{\sigma}^2] =$

- $\sigma[\hat{\sigma}] =$

Likelihood Function

- Likelihood $\mathcal{L}(x; \theta)$:
-
- With an ensemble of measurements, overall likelihood is obtained from the product of the likelihoods of the measurements
- Here θ represents one or more parameters

Log Likelihood

- To estimate the parameter(s),
maximise the likelihood
 - Set derivative to zero
 - Typically easier to maximise the
- $\frac{\partial \mathcal{L}}{\partial \theta} =$
- If there are several
we can minimise with respect to each of them

Likelihood Example: Poisson

- independent trials with results
- Likelihood function for observing if true mean is
 - $\mathcal{L}(n_i; \mu) =$
- Product over N measurements
 - $\mathcal{L}(\text{data}, \mu) =$
 - $\Rightarrow \ln \mathcal{L} =$

Best estimator is
the mean value

Likelihood Example: Gaussian

- $G(x | \mu, \sigma) =$
- Take the derivative of the log likelihood

$$\bullet \frac{\partial}{\partial \mu} (\ln \mathcal{L})|_{\hat{\mu}=\mu} =$$

- The unbiased estimator for σ is
- $\hat{\sigma} =$

Binned vs unbinned likelihood functions

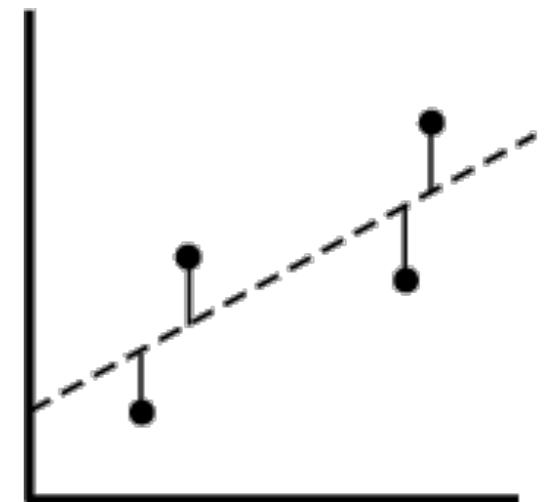
- Likelihood formalism works for any
 - Product of the $f_i(t)$ is a
 - **Example measurement:** Measure the τ of a particle of a given species for an ensemble of such particle produced at $t = 0$ such that the decay at time t :
$$\bullet f(t) = \frac{1}{\tau} e^{-t/\tau}$$
 - Consider two ways to construct the likelihood
 -
 -
 -

The Likelihood and χ^2

- If the data is Gaussian, we have
 - $\ln \mathcal{L} =$
- Compare to
 - $\chi^2 = \sum_i^N \frac{(x_i - \mu)^2}{\sigma^2}$
- By inspection
 - $\chi^2 =$
- Note: the likelihood formulation works for all pdfs not just Gaussians

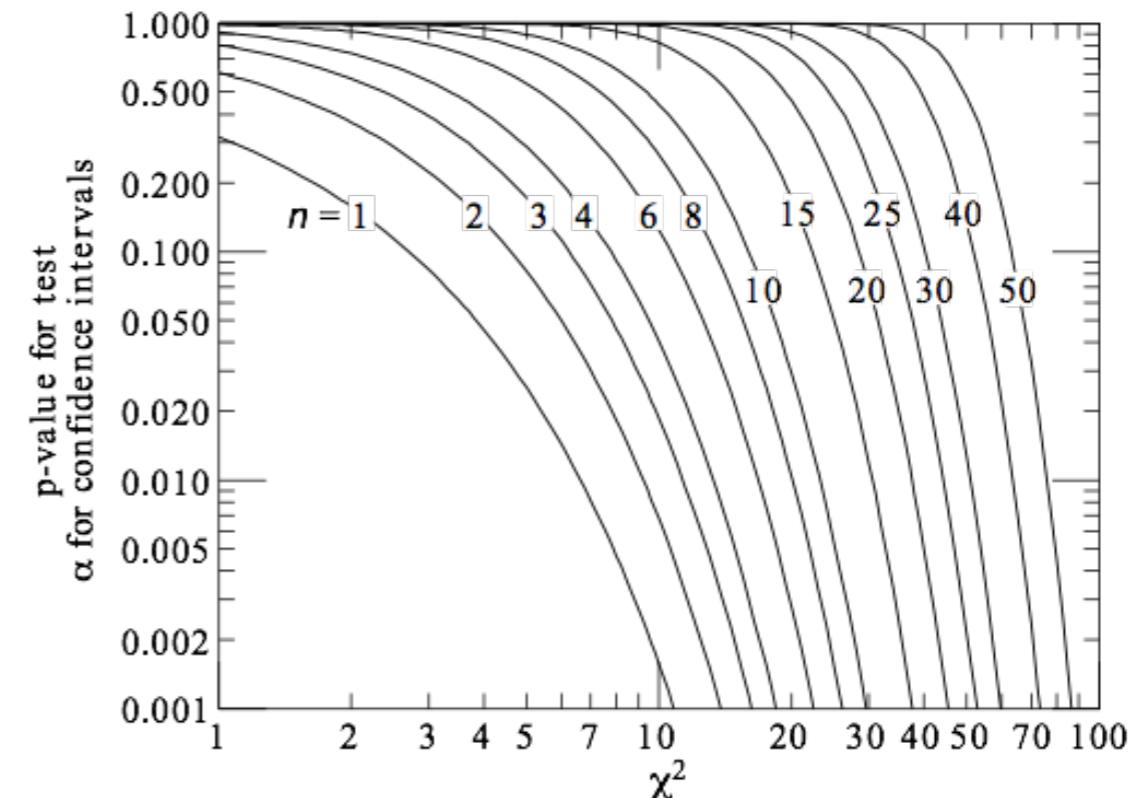
Method of Least Squares

- Assume we have enough statistics for our measurements such that we can assume we are in the Gaussian regime
- Goal:
-
- Scatter defined by χ^2
 - $\chi^2 =$
- Can write the χ^2 in terms of our observables
 - $\chi^2 =$
- Minimise χ^2 with respect to θ
- Useful when minimising $\ln \mathcal{L}$ is slow (high statistics samples)



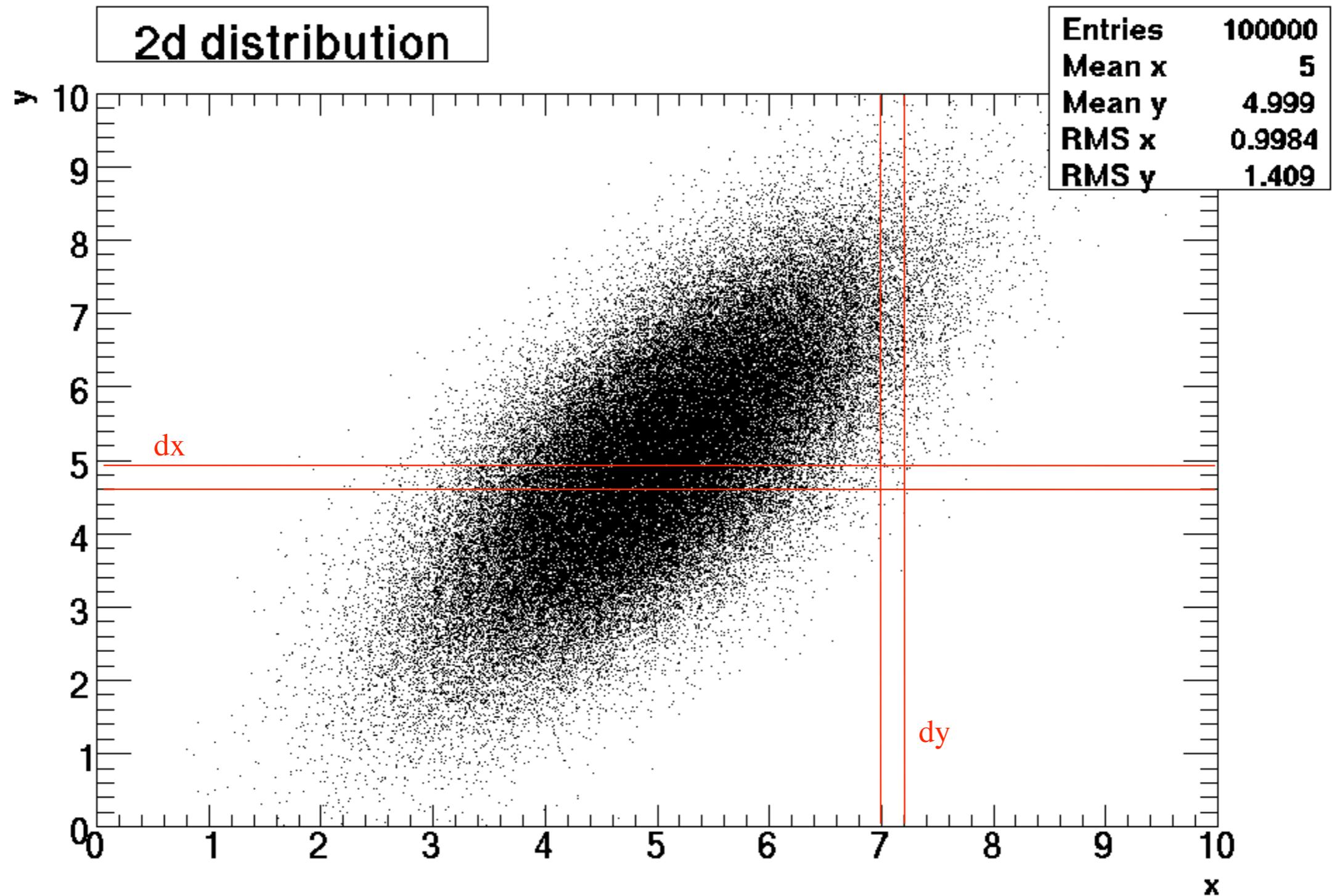
Example: chi-squared p-values

- One advantage of a is that the value of the can be interpreted as a
 - iff on each data point are and
 - the (distribution of around their are)
- In the plot below
- $n =$
- For a , expect to be close to

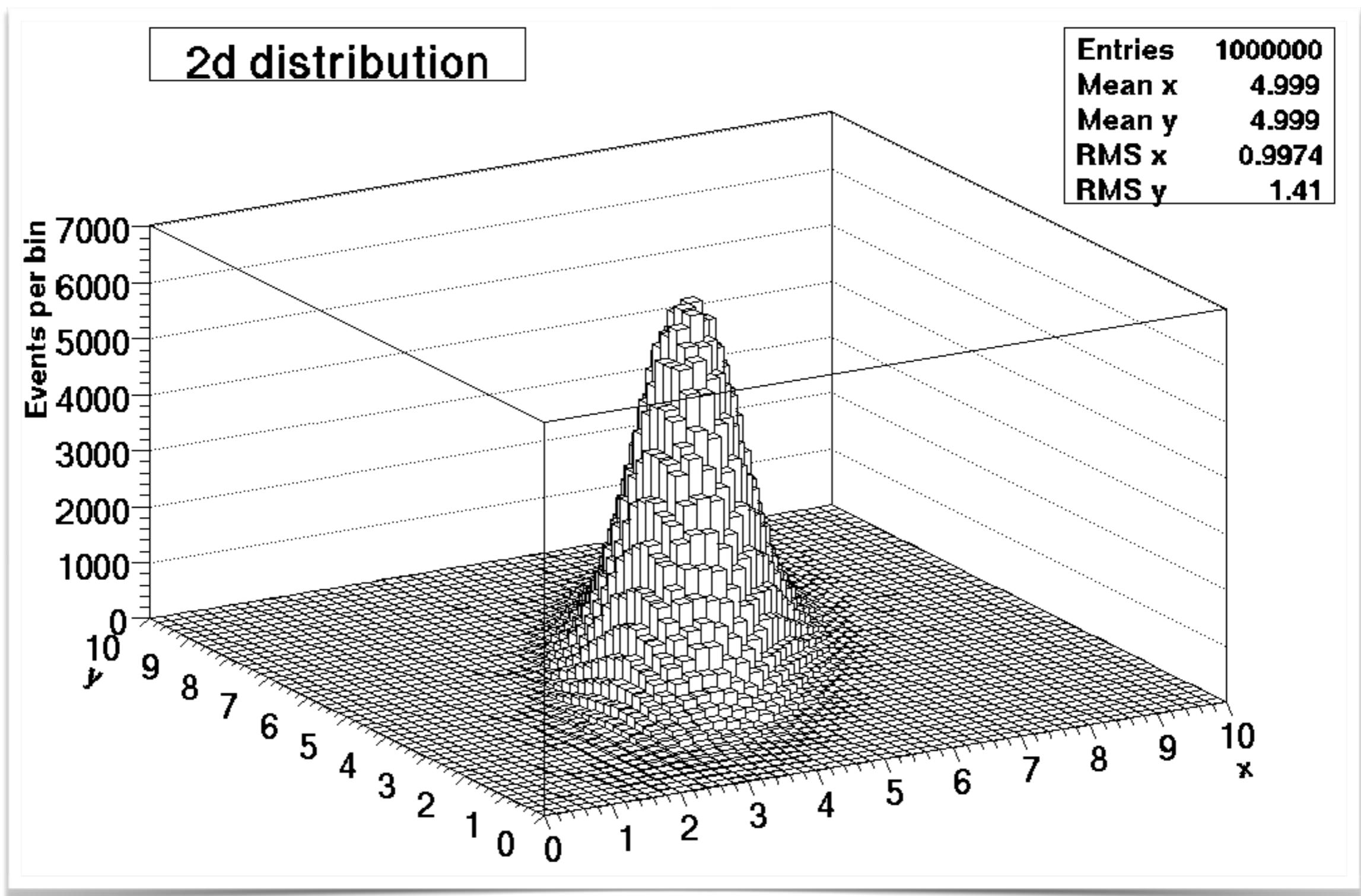


Example in jupyter notebook

2D distribution



2d distribution



Covariance and Correlation

Covariance Matrix

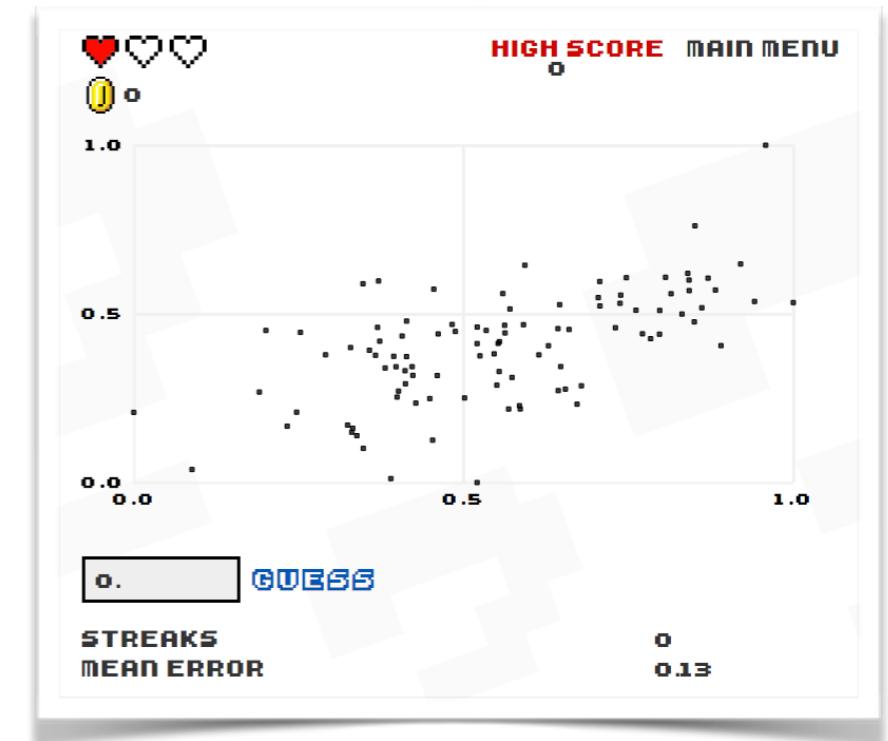
$$\text{cov}[x, y] =$$

- A representation of the N-dimensional parameter space as a covariance matrix
 - Diagonal elements:
 - Off-diagonal:

Correlation (normalised covariance)

If two variables are uncorrelated, independent variables, then $\text{cov}[x, y] = 0$ for $x \neq y$

$$\rho_{xy} =$$



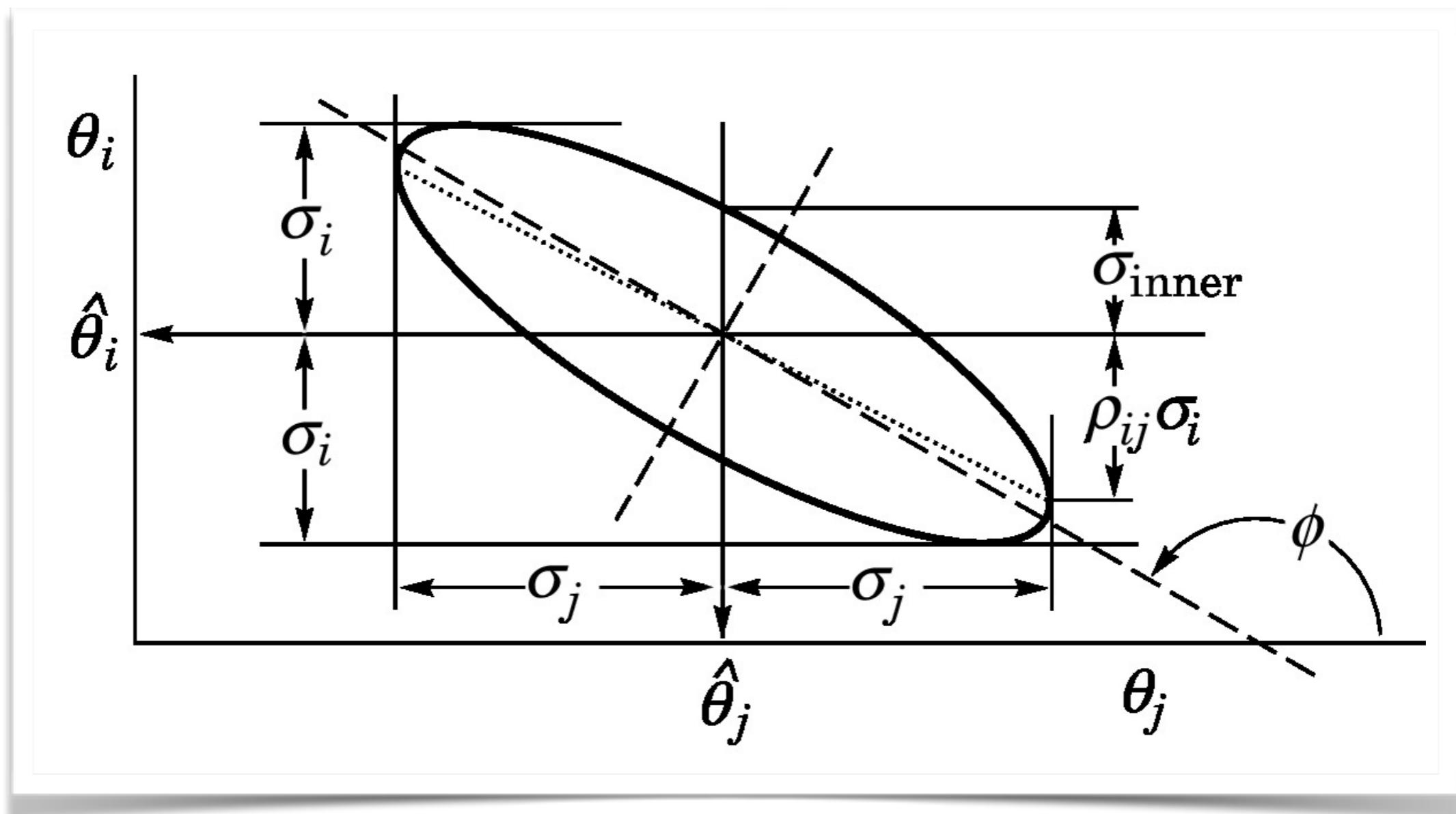
<http://guessthecorrelation.com/>

Covariance Matrix for a Gaussian

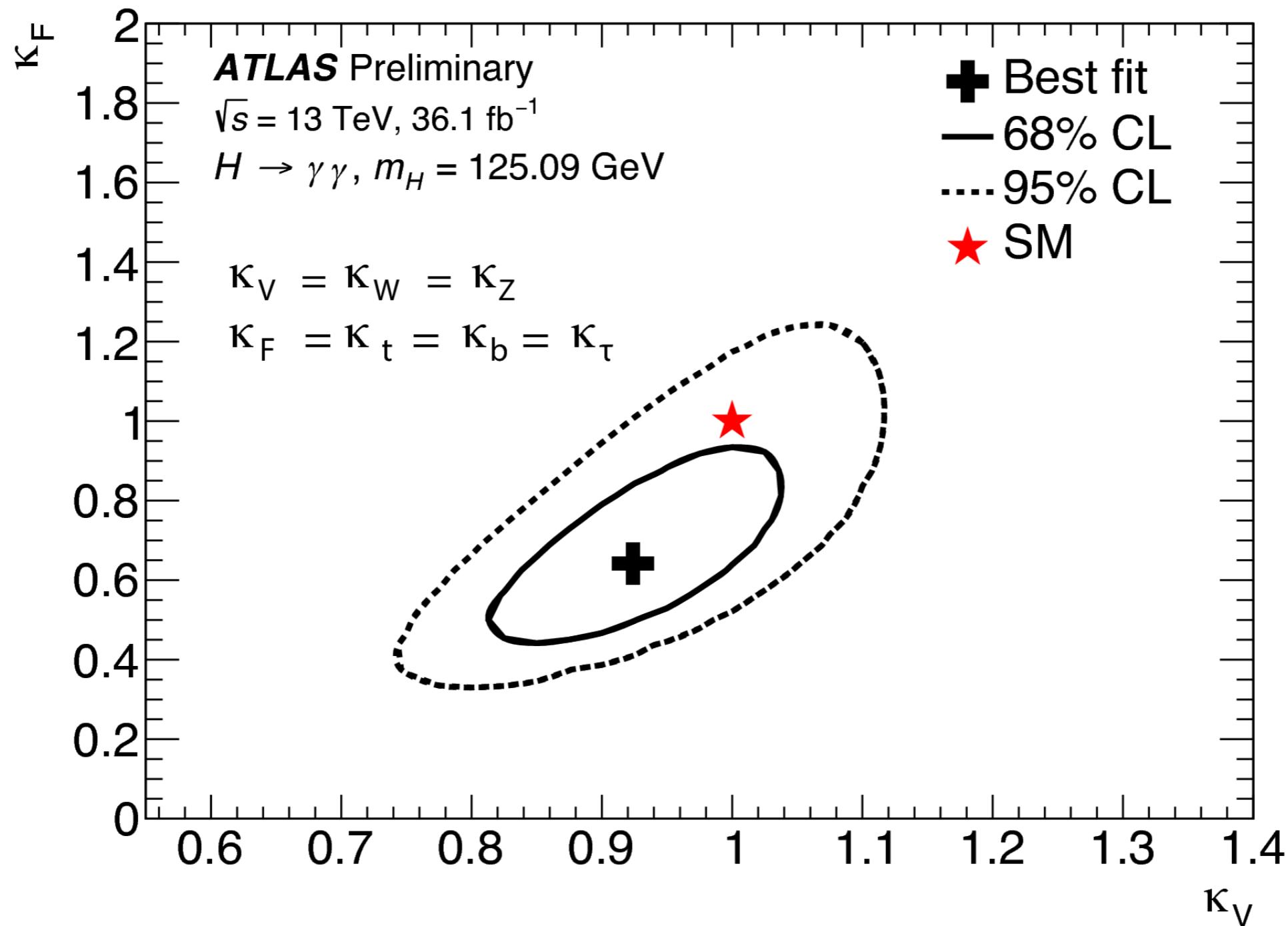
- If x and y are independent variables
 - $G(x, y | \mu_x, \sigma_x, \mu_y, \sigma_y) =$
 -
- Now, assume that x and y are correlated
- Covariance matrix is defined by
 - $\langle \hat{V}^{-1} \rangle_{ij} =$
 - For a binned likelihood, where N is large and the likelihood can be reduced to a χ^2
 - $\langle \hat{V}^{-1} \rangle_{ij} =$

Correlated Uncertainties

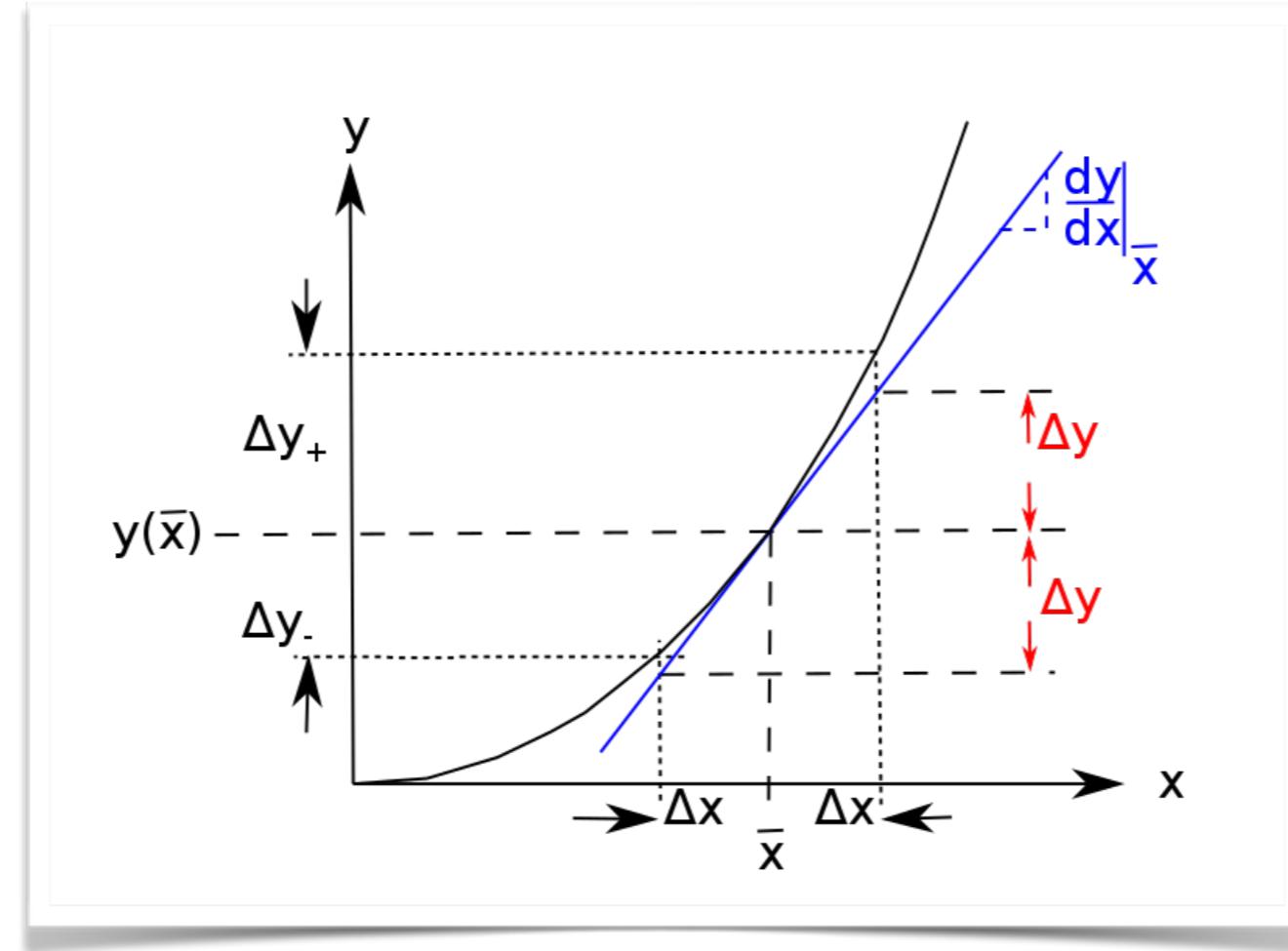
- Standard for two parameters with a correlation
- Slope related to correlation coefficient
- Correlation matrix typically determined from data numerically during fitting procedure



Correlation Example from Higgs



Propagation of Errors

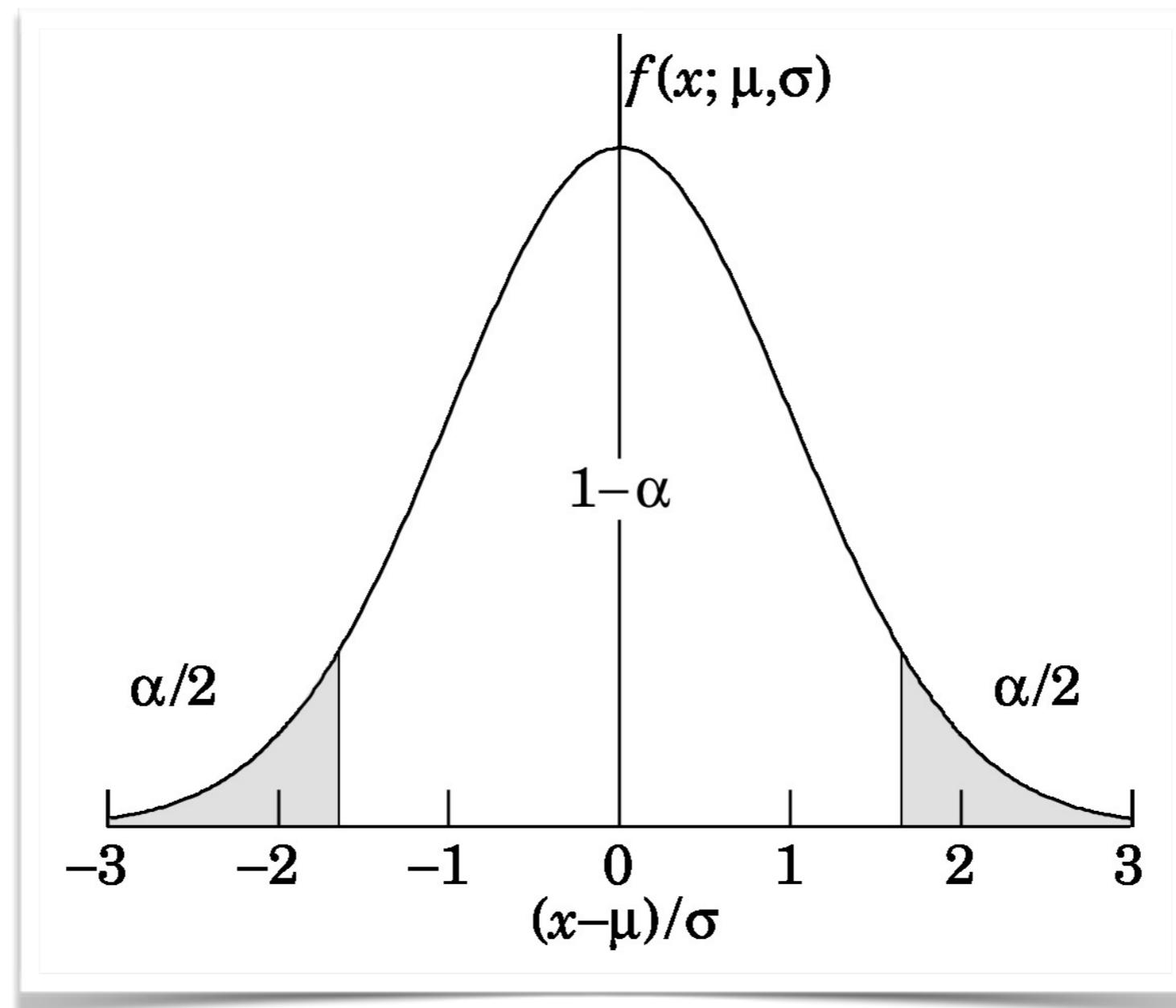


- Determine error on final measurement from known errors on input measurements
- $\sigma_f^2 =$
- More dimensions are usually expressed as a matrix
- Useful reference: https://en.wikipedia.org/wiki/Propagation_of_uncertainty

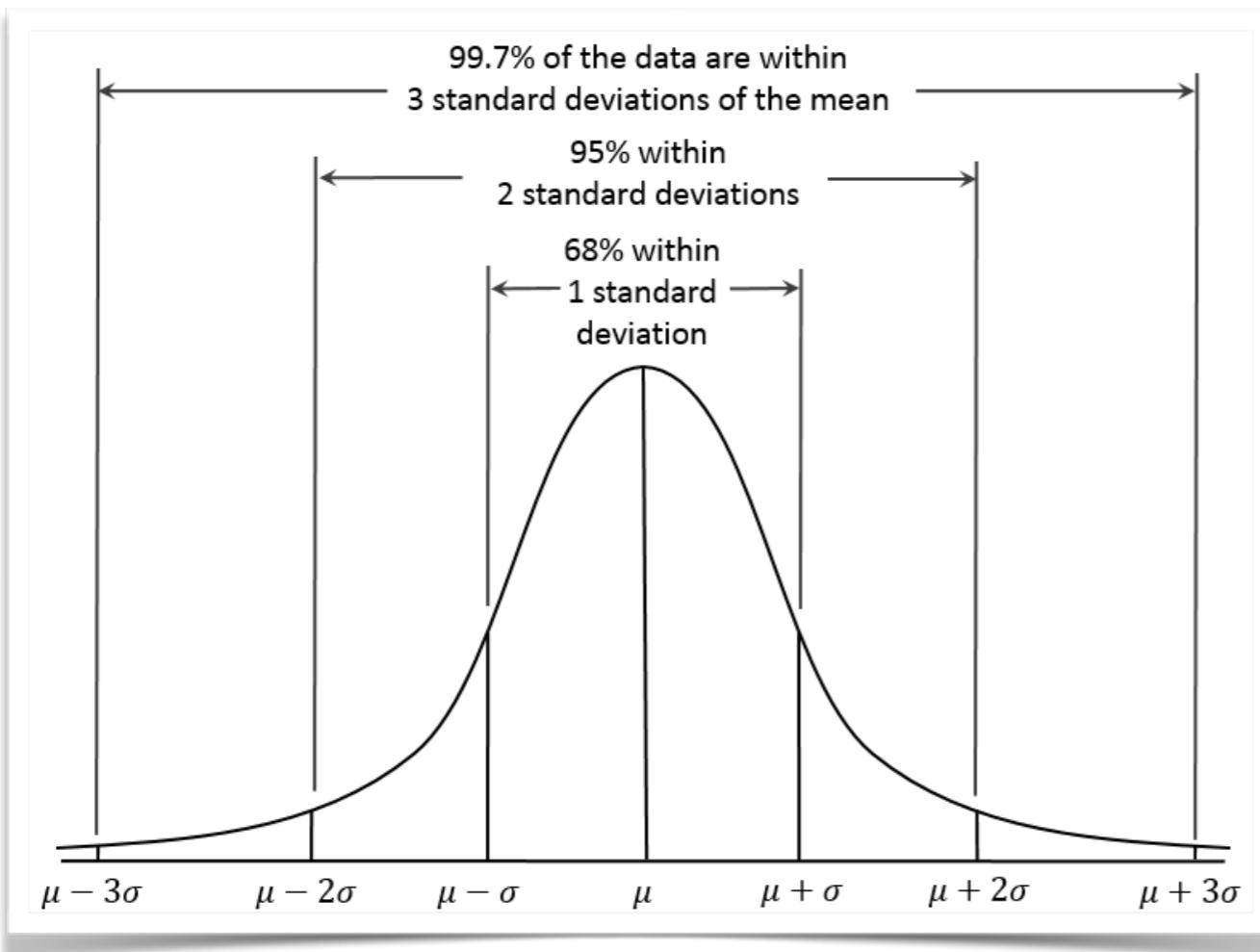
Confidence Interval

- Fraction of the result not between \underline{x} and \bar{x} is

- $1 - \alpha = \int_{\underline{x}}^{\bar{x}} f(x; \mu, \sigma) dx$



Confidence Levels for a Gaussian



α	δ
0.3173	
4.55×10^{-2}	
2.7×10^{-3}	
5.7×10^{-7}	
2.0×10^{-9}	

Confidence Levels: Higgs

