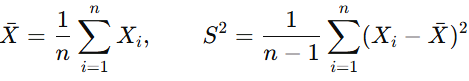
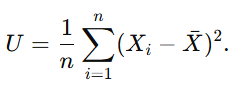
**Sampling Mean, Variance, and the Law of Large Numbers (rewritten theory)**

Let X1,…, Xn be i.i.d. with population mean μ=E[Xi] and variance σ2=Var(Xi) (finite).  
Sample statistics:

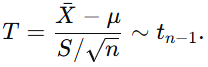
(“corrected”/Bessel’s) and the uncorrected version



**1) Properties of the sampling mean**

* **Unbiasedness**:

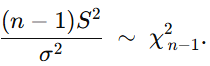


* **Variance / standard error**:, With unknown σ, replace by
* **Distribution**
  + If the parent is **normal**, then
  + **CLT (general parents)**:. Hence for large n the mean is approximately normal.
  + If parent is normal and σ unknown, the **t-statistic**



* **Consistency (LLN)**: (and almost surely under mild conditions).
* **Independence (normal case only)**: For normal parents, and S2 are independent—useful for inference.

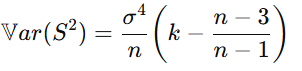
**2) Properties of the sampling variance**

* **Unbiasedness (corrected)**:.
* **Bias (uncorrected)**: (biased low).
* **Sampling distribution (normal parent)**:

Consequently,



For U,

* **General-parent variability**: letting k=μ4/σ4 be kurtosis,

****(reduces to 2σ4/(n−1) when k=3, i.e., normal). Heavy tails (k>3) inflate the dispersion of S2.

* **Consistency**:

**3) Law of Large Numbers (LLN)**

* **Weak LLN**:.
* **Strong LLN**: under similar conditions.  
  **Intuition**: averages stabilize as nn grows; variability shrinks at rate (via SE).

**4) Illustrative applications in cybersecurity**

* **Traffic baselining & anomaly detection**  
  Use rolling (e.g., bytes/flow, inter-arrival time, failed-login count). By LLN the baseline stabilizes; trigger alerts when
* **User behavior analytics**  
  Per-user averages of logins, session duration, API calls. Significant deviations from the learned mean/variance flag account takeovers or insider threats.
* **Detector performance estimation**  
  Estimate false-positive rate p from repeated trials; by LLN, and guides sample size and confidence intervals.
* **Randomness tests for crypto/keys**  
  For RNG or key-material checks, means/variances of bit streams should match theory (e.g., mean ≈0.5); chi-square and variance-based tests rely on the sampling distributions above.
* **Capacity planning / DDoS triage**  
  Stable long-run means of requests/sec help distinguish genuine step changes from short spikes; heavy-tailed metrics warn that S2 is more volatile (use robust estimators).

**5) Takeaways**

* is is unbiased, approximately/ exactly normal (CLT/normal parent), and concentrates at rate .
* S2 is unbiased with Bessel’s correction; under normality its scaled form is (Xn−1)2 with variance 2σ4/(n−1); heavy tails increase its variability.
* LLN underpins practical baselining and thresholding in security analytics, while awareness of distributional assumptions (normal vs heavy-tailed) prevents overconfident decisions.